

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

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

This is the Accepted version of the following publication

Cust, Emily, Sweeting, Alice, Ball, Kevin and Robertson, Samuel (2019)
Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *Journal of Sports Sciences*, 37 (5). pp. 568-600. ISSN 0264-0414 (In Press)

The publisher's official version can be found at
<https://www.tandfonline.com/doi/full/10.1080/02640414.2018.1521769>
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/37960/>

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:	<p>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).</p>
Order of Authors:	Emily Elizabeth Cust Alice J Sweeting Kevin Ball Sam Robertson
Response to Reviewers:	<p>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.</p> <p>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.</p> <p>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.</p> <p>Reviewer #3:</p>

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 excluded. 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

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

3
4 *Emily E. Cust^{1, 2*}, Alice J. Sweeting^{1, 2}, Kevin Ball¹ and Sam Robertson^{1, 2}*

5
6 **Author details:**

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

8 ² Western Bulldogs Football Club, Footscray, Melbourne, Australia

9
10 **Author ORCID**

11 Sam Robertson 0000-0002-8330-0011

12 Alice Sweeting 0000-0002-9185-6773

13 Emily Cust 0000-0001-6927-6329

14
15 **Author contact details:**

16 *** Corresponding author:**

17 Emily E. Cust

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

19 Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001,
20 Australia. Western Bulldogs Football Club, Footscray, Melbourne, Australia

21
22 Alice J. Sweeting

23 Email: Alice.Sweeting@vu.edu.au

24 Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001,
25 Australia. Western Bulldogs Football Club, Footscray, Melbourne, Australia

26
27 Kevin Ball

28 Email: Kevin.Ball@vu.edu.au

29 Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001,
30 Australia.

31
32 Sam Robertson

33 Email: Sam.Robertson@vu.edu.au

34 Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001,
35 Australia. Western Bulldogs Football Club, Footscray, Melbourne, Australia

36
37 **Running title:**

38 Machine and deep learning for sport movement recognition review

39
40 .

41 **Abstract**

42

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43 Objective assessment of an athlete’s performance is of importance in elite sports to facilitate
44 detailed analysis. The implementation of automated detection and recognition of sport-specific
45 movements overcomes the limitations associated with manual performance analysis methods. The
46 object of this study was to systematically review the literature on machine and deep learning for
47 sport-specific movement recognition using inertial measurement unit (IMU) and, or computer
48 vision data inputs. A search of multiple databases was undertaken. Included studies must have
49 investigated a sport-specific movement and analysed via machine or deep learning methods for
50 model development. A total of 52 studies met the inclusion and exclusion criteria. Data pre-
51 processing, processing, model development and evaluation methods varied across the studies.
52 Model development for movement recognition were predominantly undertaken using supervised
53 classification approaches. A kernel form of the Support Vector Machine algorithm was used in
54 53% of IMU and 50% of vision-based studies. Twelve studies used a deep learning method as a
55 form of Convolutional Neural Network algorithm and one study also adopted a Long Short Term
56 Memory architecture in their model. The adaptation of experimental set-up, data pre-processing,
57 and model development methods are best considered in relation to the characteristics of the
58 targeted sports movement(s).

59

60

61 **Key Words:**

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

65 1. Introduction

66

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

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

94 and tackling impacts in Australian Rules football (Gastin et al., 2014; Gastin, McLean, Spittle, &
95 Breed, 2013) and rugby (Gabbett, Jenkins, & Abernethy, 2012, 2011; Howe, Aughey, Hopkins,
96 Stewart, & Cavanagh, 2017; Hulin, Gabbett, Johnston, & Jenkins, 2017). Other applications
97 include swimming analysis (Mooney, Corley, Godfrey, Quinlan, & ÓLaighin, 2015), golf swing
98 kinematics (Lai, Hetchl, Wei, Ball, & McLaughlin, 2011), over-ground running speeds (Wixted,
99 Billing, & James, 2010), full motions in alpine skiing (Yu et al., 2016); and the detection and
100 evaluation of cricket bowling (McNamara, Gabbett, Blanch, & Kelly, 2017; McNamara, Gabbett,
101 Chapman, Naughton, & Farhart, 2015; Wixted, Portus, Spratford, & James, 2011).

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

119 Machine learning algorithms learn from data input for automated model building and
120 perform tasks without being explicitly programmed. The algorithm goal is to output a response
121 function $\overline{f_{\sigma}(\bar{x})}$ that will predict a ground truth variable \bar{y} from an input vector of variables \bar{x} . Models
122 are run for classification techniques to predict a target class (Kotsiantis, Zaharakis, & Pintelas,
123 2007), or regression to predict discrete or continuous values. Models are aimed at finding an

124 optimal set of parameters $\bar{\varphi}$ to describe the response function, and then make predictions on unseen
125 unlabelled data input. Within these, model training approaches can generally run as supervised
126 learning, unsupervised learning or semi-supervised learning (Mohammed, Khan, & Bashier, 2016;
127 Sze, Chen, Yang, & Emer, 2017).

128 Processing raw data is limited for conventional machine learning algorithms, as they are
129 unable to effectively be trained on abstract and high-dimensional data that is inconsistent, contains
130 missing values or noisy artefacts (Bux et al., 2017; Kautz, 2017). Consequently, several pre-
131 processing stages are required to create a suitable data form for input into the classifier algorithm
132 (Figo, Diniz, Ferreira, & Cardoso, 2010). Filtering (Figo et al., 2010; Wundersitz, Gastin,
133 Robertson, Davey, & Netto, 2015), window capture durations (Mitchell, Monaghan, & O'Connor,
134 2013; Preece, Goulermas, Kenney, & Howard, 2009; Wundersitz, Josman, et al., 2015), and signal
135 frequency cut-offs (Wundersitz, Gastin, Richter, Robertson, & Netto, 2015; Wundersitz, Gastin,
136 Robertson, et al., 2015) are common techniques applied prior to data prior to dynamic human
137 movement recognition. Well-established filters for processing motion signal data include the
138 Kalman filter (Kautz, 2017; Titterton & Weston, 2009; D. Wagner, Kalischewski, Velten, &
139 Kummert, 2017) and a Fourier transform filter (Preece, Goulermas, Kenney, Howard, et al., 2009)
140 such as a fast Fourier transform (Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015;
141 Preece, Goulermas, Kenney, & Howard, 2009). Near real-time processing benefits from reducing
142 memory requirements, computational demands, and essential bandwidth during whole model
143 implementation. Signal feature extraction and selection favours faster processing by reducing the
144 signals to the critical features that can discriminate the targeted activities (Bulling et al., 2014).
145 Feature extraction involves identifying the key features that help maximise classifier success, and
146 removing features that have minimal impact in the model (Mannini & Sabatini, 2010). Thus,
147 feature selection involves constructing data representations in subspaces with reduced dimensions.
148 These identified variables are represented in a compact feature variable (Mannini & Sabatini,
149 2010). Common methods include principal component analysis (PCA) (Gløersen, Myklebust,
150 Hallén, & Federolf, 2018; Young & Reinkensmeyer, 2014), vector coding techniques (Hafer &
151 Boyer, 2017) and empirical cumulative distribution functions (ECDF) (Plötz, Hammerla, &
152 Olivier, 2011). An ECDF approach has been shown to be advantageous over PCA as it derives
153 representations of raw input independent of the absolute data ranges, whereas PCA is known to

154 have reduced performance when the input data is not properly normalised (Plötz et al., 2011). For
155 further detailed information on the acquisition, filtering and analysis of IMU data for sports
156 application and vision-based human activity recognition, see (Kautz, 2017) and (Bux et al., 2017),
157 respectively.

158 Deep learning is a division of machine learning, characterised by deeper neural network
159 model architectures and are inspired by the biological neural networks of the human brain (Bengio,
160 2013; LeCun et al., 2015; Sze et al., 2017). The deeper hierarchical models create a profound
161 architecture of multiple hidden layers based on representative learning with several processing and
162 abstraction layers (Bux et al., 2017; J. B. Yang et al., 2015). These computational models allow
163 data input features to be automatically extracted from raw data and transformed to handle
164 unstructured data, including vision (LeCun et al., 2015; Ravi, Wong, Lo, & Yang, 2016). This
165 direct input avoids several processing steps required in machine learning during training and
166 testing, therefore reducing overall computational times. A current key element within deep learning
167 is backpropagation (Hecht-Nielsen, 1989; LeCun, Bottou, Orr, & Müller, 1998). Backpropagation
168 is a fast and computationally efficient algorithm, using gradient descent, that allows training deep
169 neural networks to be tractable (Sze et al., 2017). Human activity recognition has mainly been
170 performed using conventional machine learning classifiers. Recently, deep learning techniques
171 have enhanced the bench mark and applications for IMUs (Kautz et al., 2017; Ravi et al., 2016;
172 Ronao & Cho, 2016; J. B. Yang et al., 2015; Zebin, Scully, & Ozanyan, 2016; Zeng et al., 2014)
173 and vision (Ji, Yang, Yu, & Xu, 2013; Karpathy et al., 2014a; Krizhevsky, Sutskever, & Hinton,
174 2012; Nibali, He, Morgan, & Greenwood, 2017) in human movement recognition producing more
175 superior model performance accuracy.

176 The objective of this study was to systematically review the literature investigating sport-
177 specific automated movement detection and recognition. The review focusses on the various
178 technologies, analysis techniques and performance outcome measures utilised. There are several
179 reviews within this field that are sensor-based including wearable IMUs for lower limb
180 biomechanics and exercises (Fong & Chan, 2010; M. O'Reilly, Caulfield, Ward, Johnston, &
181 Doherty, 2018), swimming analysis (Magalhaes, Vannozzi, Gatta, & Fantozzi, 2015; Mooney et
182 al., 2015), quantifying sporting movements (Chambers et al., 2015) and physical activity
183 monitoring (C. C. Yang & Hsu, 2010). A recent systematic review has provided an evaluation on

184 the in-field use of inertial-based sensors for various performance evaluation applications
185 (Camomilla et al., 2018). Vision-based methods for human activity recognition (Aggarwal & Xia,
186 2014; Bux et al., 2017; Ke et al., 2013; Zhang et al., 2017), semantic human activity recognition
187 (Ziaeefard & Bergevin, 2015) and motion analysis in sport (Barris & Button, 2008) have also been
188 reviewed. However, to date, there is no systematic review across sport-specific movement
189 detection and recognition via machine or deep learning. Specifically, incorporating IMUs and
190 vision-based data input, focussing on in-field applications as opposed to laboratory-based protocols
191 and detailing the analysis and machine learning methods used.

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

199

200 **2. Methods**

201

202 **2.1 Search strategy**

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

213

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

215

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

217

218 **2.2 Data extraction**

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

244 results or information identified from the studies was included as a special remark (Sprager &
245 Juric, 2015). Hardware and specification information extracted included the IMU or video
246 equipment used, number of units, attachment of sensors (IMUs), sample frequency, and sensor data
247 types used in analysis (IMUs). Studies identified and full data extracted were reviewed by a second
248 author.

250 **3. Results**

251
252 An outline of the search results and study exclusions has been provided in Fig 1. Of the initial
253 database search which identified 4885 results, a final 52 studies met criteria for inclusion in this
254 review. Of these, 29 used IMUs and 22 were vision-based. One study (Ó Conaire et al., 2010) used
255 both sensors and vision for model development separately then together via data fusion. Tables 3 -
256 8 provide a description of the characteristics of the reviewed studies, detailed in the following
257 sections.

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

261 **3.1 Experimental design**

262 A variety of sports and their associated sport-specific movements were investigated, implementing
263 various experimental designs as presented in Tables 5 and 7. Across the studies, sports reported
264 were tennis (n = 10), cricket (n = 3), weightlifting or strength training (n = 6), swimming (n = 4),
265 skateboarding (n = 2), ski jumping (n = 2), snowboarding (n = 1), golf (n = 4), volleyball (n = 2),
266 rugby (n = 2), ice hockey (n = 2), gymnastics (n = 2), karate (n = 1), basketball (n = 3), Gaelic
267 football (n = 1), hurling (n = 1), boxing (n = 2), running (n = 2), diving (n = 1), squash (n = 1),
268 badminton (n = 1), cross-country skiing (n = 2) and soccer (n = 4). The Sports 1-M dataset
269 (Karpathy et al., 2014b) was also reported, which consists of 1,133,158 video URLs annotated
270 automatically with 487 sport labels using the YouTube Topic API. A dominant approach was the
271 classification of main characterising actions for each sport. For example, serve, forehand, backhand
272 strokes in tennis (Connaghan et al., 2011; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Shah,
273 Chokalingam, Paluri, & Pradeep, 2007; Srivastava et al., 2015), and the four competition strokes in

274 swimming (Jensen, Blank, Kugler, & Eskofier, 2016; Jensen, Prade, & Eskofier, 2013; Liao et al.,
275 2003; Victor et al., 2017). Several studies further classified sub-categories of actions. For example,
276 three further classes of the two main classified snowboarding trick types Grinds and Airs (Groh,
277 Fleckenstein, & Eskofier, 2016), and further classifying the main tennis stroke types as either flat,
278 topspin or slice (Srivastava et al., 2015). Semantic descriptors were reported for classification
279 models that predicted athlete training background, experience and fatigue level. These included
280 running (Buckley et al., 2017; Kobsar, Osis, Hettinga, & Ferber, 2014), rating of gymnastic
281 routines (Reily, Zhang, & Hoff, 2017), soccer pass classification based on its quality (Horton,
282 Gudmundsson, Chawla, & Estephan, 2014), cricket bowling legality (Qaisar et al., 2013; Salman,
283 Qaisar, & Qamar, 2017), ski jump error analysis (Brock & Ohgi, 2017; Brock, Ohgi, & Lee, 2017)
284 and strength training technique deviations (M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield,
285 2017a; M. O'Reilly et al., 2015; M. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017). One
286 approach (Yao & Fei-Fei, 2010), encoded the mutual context of human pose and sporting
287 equipment using semantics, to facilitate the detection and classification of movements including a
288 cricket bat and batsman coupled movements.

289 Total participant numbers for IMU-based studies ranged from one (Qaisar et al., 2013) to
290 30 (Kautz et al., 2017). Reported data individual instance sample sizes for sensor studies ranged
291 from 150 (Salman et al., 2017) to 416, 737 (Rassem, El-Beltagy, & Saleh, 2017). Vision-based
292 studies that explicitly reported total participant details ranged from five (Ó Conaire et al., 2010) to
293 40 (Victor et al., 2017). Vision dataset sample sizes varied across studies, from 50 individual action
294 clips (Liao et al., 2003) to 15, 000 (Victor et al., 2017). One study (Karpathy et al., 2014a) used the
295 publicly available Sports-1M, as previously described. Vision-based studies also reported datasets
296 in total time, 10.3 hours (Bertasiu, Park, Yu, & Shi, 2017), 3 hours (Montoliu, Martín-Félez,
297 Torres-Sospedra, & Martínez-Usó, 2015), 1, 500 minutes (Shah et al., 2007), and 50 hours (Kapela
298 et al., 2015), and by frame numbers, 6, 035 frames (Zhu, Xu, Gao, & Huang, 2006) and 10, 115
299 frames (Reily et al., 2017).

300

301 **3.2 Inertial measurement unit specifications**

302 A range of commercially available and custom-built IMUs were used in the IMU-based studies (n=
303 30), as presented in Table 3. Of these, 23% reported using a custom-built sensor. Of the IMU-based

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

304 studies, the number of sensors mounted or attached to each participant or sporting equipment piece
305 ranged from one to nine. The majority of studies (n= 22) provided adequate details of sensor
306 specifications including sensor type, axes, measurement range, and sample rate used. At least one
307 characteristic of sensor measurement range or sample rate used in data collection was missing from
308 eight studies. All studies used triaxial sensors and collected accelerometer data. For analysis and
309 model development, individual sensor data consisted of only accelerometer data (n = 8), both
310 accelerometer and gyroscope data (n = 15), and accelerometer, gyroscope and magnetometer data
311 (n = 7). The individual sensor measurement ranges reported for accelerometer were ± 1.5 g to ± 16
312 g, gyroscope ± 500 °/s to ± 2000 °/s, magnetometer ± 1200 μ T or 1.2 to 4 Ga. Individual sensor
313 sample rates ranged from 10 Hz to 1000 Hz for accelerometers, 10 Hz to 500 Hz for gyroscopes
314 and 50 Hz to 500 Hz for magnetometers.

315

316 *** Table 3 near here***

317

318 3.3 Vision capture specification

319 Several experimental set-ups and specifications were reported in the total 23 vision-based studies
320 (Table 4). Modality was predominately red, green, blue (RGB) cameras. Depth cameras were
321 utilised (Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015; Kasiri, Fookes, Sridharan,
322 & Morgan, 2017; Reily et al., 2017), which add depth perception for 3-dimensional image
323 mapping. Seven studies clearly reported the use of a single camera set-up (Couceiro, Dias, Mendes,
324 & Araújo, 2013; Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014; Hachaj, Ogiela, &
325 Koptyra, 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Nibali et al., 2017; Reily et al.,
326 2017). One study reported 16 stationary positioned cameras at a ‘bird’s eye view’ (Montoliu et al.,
327 2015), and Ó Conaire et al. (2010) reported the use of one overhead and 8 stationary cameras
328 around a tennis court baseline, although data from two cameras were only used in final analysis due
329 to occlusion issues. Sample frequency and, or pixel resolution were reported in seven of the studies
330 (Couceiro et al., 2013; Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017;
331 Montoliu et al., 2015; Victor et al., 2017; Zhu et al., 2006), with sample frequencies ranging from
332 30 Hz to 210 Hz.

333

334 *** Table 4 near here***

335

336 3.4 Inertial measurement unit recognition model development methods

337 Key stages of model development from data pre-processing to recognition techniques for IMU-
338 based studies are presented in Table 5. Data pre-processing filters were reported as either a low-
339 pass filter (n = 7) (Adelsberger & Tröster, 2013; Buckley et al., 2017; Kelly, Coughlan, Green, &
340 Caulfield, 2012; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2015, 2017; Rindal, Seeberg,
341 Tjønnås, Haugnes, & Sandbakk, 2018), high-pass filter (n = 2) (Kautz et al., 2017; Schuldhaus et
342 al., 2015), or calibration with a filter (Salman et al., 2017). Processing methods were reported in
343 67% of the IMU-based studies (Adelsberger & Tröster, 2013; Anand, Sharma, Srivastava,
344 Kaligounder, & Prakash, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe, Blanke, Capkevics,
345 & Tröster, 2016; Groh et al., 2016; Groh, Fleckenstein, Kautz, & Eskofier, 2017; Groh, Kautz, &
346 Schuldhaus, 2015; Jensen et al., 2016, 2015; Jiao, Wu, Bie, Umek, & Kos, 2018; Kautz et al.,
347 2017; Kobsar et al., 2014; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Ó Conaire et al.,
348 2010; Pernek, Kurillo, Stiglic, & Bajcsy, 2015; Qaisar et al., 2013; Salman et al., 2017; Schuldhaus
349 et al., 2015). Methods included, calibration of data (Groh et al., 2016, 2017; Jensen et al., 2015;
350 Qaisar et al., 2013), a one-second window centred around identified activity peaks in the signal
351 (Adelsberger & Tröster, 2013; Schuldhaus et al., 2015), temporal alignment (Pernek et al., 2015),
352 normalisation (Ó Conaire et al., 2010), outlier adjustment (Kobsar et al., 2014) or removal (Salman
353 et al., 2017), and sliding windows ranging from one to 3.5 seconds across the data (Jensen et al.,
354 2016). The three studies that investigated trick classification in skateboarding (Groh et al., 2017,
355 2015) and snowboarding (Groh et al., 2016) corrected data for different rider board stance styles,
356 termed Regular or Goofy, by inverting signal axes.

357 Movement detection methods were specifically reported in 16 studies (Adelsberger &
358 Tröster, 2013; Anand et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015, Jensen et
359 al., 2013, 2015; Kautz et al., 2017; Kelly et al., 2012; Kos & Kramberger, 2017; Ó Conaire et al.,
360 2010; Rindal et al., 2018; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside, Cant, Connolly,
361 & Reid, 2017). Detection methods included thresholding (n = 5), windowing segmenting (n = 4),
362 and a combination of threshold and windowing techniques (n = 5).

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

368 Algorithms trialled for movement recognition were diverse across the literature (Table 5).
369 Supervised classification using a kernel form of Support Vector Machine (SVM) was most
370 prevalent (n = 16) (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Brock et al., 2017; Buckley
371 et al., 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Kautz et al.,
372 2017; Kelly et al., 2012; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017;
373 Schuldhaus et al., 2015; Whiteside et al., 2017). The next highest tested were Naïve Bayesian (NB)
374 (n = 8) (Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015; Kautz et al.,
375 2017; Salman et al., 2017; Schuldhaus et al., 2015) and k-Nearest Neighbour (kNN) (n = 8)
376 (Buckley et al., 2017; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Ó Conaire et al., 2010;
377 Salman et al., 2017; Whiteside et al., 2017), followed by Random Forests (RF) (n = 7) (Buckley et
378 al., 2017; Groh et al., 2017; Kautz et al., 2017; M. A. O'Reilly et al., 2017a; M. O'Reilly et al.,
379 2017; Salman et al., 2017; Whiteside et al., 2017). Supervised learning algorithms were the most
380 common (n = 29). One study used an unsupervised discriminative analysis approach for detection
381 and classification of tennis strokes (Kos & Kramberger, 2017). Five IMU-based study investigated
382 a deep learning approach including using Convolutional Neural Networks (CNN) (Anand et al.,
383 2017; Brock et al., 2017; Jiao et al., 2018; Kautz et al., 2017; Rassem et al., 2017) and Long Short
384 Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) architectures (Rassem et al., 2017;
385 Sharma, Srivastava, Anand, Prakash, & Kaligounder, 2017). In order to assess the effectiveness of
386 the various classifiers from each study, model performance measures quantify and visualise the
387 predictive performance as reported in the following section.

388

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

390

391 **3.5 Inertial measurement unit recognition model evaluation**

392 Reported performance evaluations of developed models across the IMU-based studies are shown in
393 Table 6. Classification accuracy, as a percentage score for the number of correct predictions by
394 total number of predictions made, was the main model evaluation measure ($n = 24$). Classification
395 accuracies across studies ranged between 52% (Brock & Ohgi, 2017) to 100% (Buckley et al.,
396 2017). Generally, the reported highest accuracy for a specific movement was $\geq 90\%$ ($n = 17$)
397 (Adelsberger & Tröster, 2013; Anand et al., 2017; Buckley et al., 2017; Connaghan et al., 2011;
398 Groh et al., 2015; Jensen et al., 2013; Jiao et al., 2018; Kobsar et al., 2014; Kos & Kramberger,
399 2017; M. A. O'Reilly et al., 2017a; Ó Conaire et al., 2010; Pernek et al., 2015; Qaisar et al., 2013;
400 Rindal et al., 2018; Schuldhaus et al., 2015; Srivastava et al., 2015; Whiteside et al., 2017) and \geq
401 80% to 90% ($n = 7$) (Brock & Ohgi, 2017; Brock et al., 2017; Groh et al., 2017; Jensen et al., 2016;
402 M. O'Reilly et al., 2015, 2017; Salman et al., 2017). As an estimate of the generalised performance
403 of a trained model on $\overline{n-x}$ samples, a form of leave-one-out cross validation (LOO-CV) was used in
404 47% of studies (Buthe et al., 2016; Groh et al., 2016, 2017, 2015, Jensen et al., 2016, 2013; Kobsar
405 et al., 2014; M. O'Reilly et al., 2015, 2017; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et
406 al., 2017; Schuldhaus et al., 2015). Precision, specificity and sensitivity (also referred to as recall)
407 evaluations were derived for detection ($n = 6$) and classification models ($n = 10$). Visualisation of
408 prediction results in the form of a confusion matrix featured in six studies (Buthe et al., 2016; Groh
409 et al., 2017; Kautz et al., 2017; Pernek et al., 2015; Rindal et al., 2018; Whiteside et al., 2017).

410

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

412

413 **3.6 Vision recognition model development methods**

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

418 Both machine ($n = 16$) and deep learning ($n = 7$) algorithms were used to recognise
419 movements from vision data. Of these, a kernel form of the SVM algorithm was most common in
420 the studies ($n = 10$) (Couceiro et al., 2013; Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri

421 et al., 2017; Li et al., 2018; Montoliu et al., 2015; M. A. O'Reilly, Whelan, Ward, Delahunt, &
422 Caulfield, 2017b; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006).
423 Other algorithms included kNN (n = 3) (Díaz-Pereira et al., 2014; Montoliu et al., 2015; Ó Conaire
424 et al., 2010), decision tree (DT) (n = 2) (Kapela et al., 2015; Liao et al., 2003), RF (n = 2) (Kasiri-
425 Bidhendi et al., 2015; Kasiri et al., 2017), and Multilayer Perceptron (MLP) (n = 2) (Kapela et al.,
426 2015; Montoliu et al., 2015). Deep learning was investigated in seven studies (Bertasius et al.,
427 2017; Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016; Karpathy et al., 2014a; Nibali et al.,
428 2017; Ramanathan et al., 2015; Tora, Chen, & Little, 2017; Victor et al., 2017) of which used
429 CNNs or LSTM RNNs as the core model structure.

430

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

432

433 **3.7 Vision recognition model evaluation**

434 Performance evaluation methods and results for vision-based studies are reported in Table 8. As
435 with IMU-based studies, classification accuracy was the common method for model evaluations,
436 featured in 61%. Classification accuracies were reported between 60.9% (Karpathy et al., 2014a)
437 and 100% (Hachaj et al., 2015; Nibali et al., 2017). In grouping the reported highest accuracies for
438 a specific movement that were $\geq 90\%$ (n = 9) (Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015;
439 Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; Nibali et al., 2017; Ó Conaire et al., 2010;
440 Reily et al., 2017; Shah et al., 2007), and $\geq 80\%$ to 90% (n = 2) (Horton et al., 2014; Yao & Fei-
441 Fei, 2010). A confusion matrix as a visualisation of model prediction results was used in nine
442 studies (Couceiro et al., 2013; Hachaj et al., 2015; Ibrahim et al., 2016; Karpathy et al., 2014a;
443 Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Lu, Okuma, & Little, 2009; Shah et al., 2007; Tora
444 et al., 2017). Two studies assessed and reported their model computational average speed (Lu et al.,
445 2009) and time (Reily et al., 2017).

446

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

448

449 **4 Discussion**

450

451 The aim of this systematic review was to evaluate the use of machine and deep learning for sport-
1
2 452 specific movement recognition from IMUs and, or computer vision data inputs. Overall, the search
3
4 453 yielded 52 studies, categorised as 29 which used IMUs, 22 vision-based and one study using both
5
6 454 IMUs and vision. Automation or semi-automated sport movement recognition models working in
7
8 455 near-real time is of particular interest to avoid the error, cost and time associated with manual
9
10 456 methods. Evident in the literature, models are trending towards the potential to provide optimised
11
12 457 objective assessments of athletic movement for technical and tactical evaluations. The majority of
13
14 458 studies achieved favourable movement recognition results for the main characterising actions of a
15
16 459 sport, with several studies exploring further applications such as an automated skill quality
17
18 460 evaluation or judgement scoring, for example automated ski jump error evaluation (Brock et al.,
19
20 461 2017).

21
22
23
24 462 Experimental set-up of IMU placement and numbers assigned per participant varied
25
26 463 between sporting actions. The sensor attachment locations set by researchers appeared dependent
27
28 464 upon the specific sporting conditions and movements, presumably to gain optimal signal data.
29
30 465 Proper fixation and alignment of the sensor axes with limb anatomical axes is important in
31
32 466 reducing signal error (Fong & Chan, 2010). The attachment site hence requires a biomechanical
33
34 467 basis for accuracy of the movement being targeted to obtain reliable data. Single or multiple sensor
35
36 468 use per person also impacts model development trade-off between accuracy, analysis complexity,
37
38 469 and computational speed or demands. In tennis studies, specificity whilst using a single sensor was
39
40 470 demonstrated by mounting the IMU on the wrist or forearm of the racquet arm (Connaghan et al.,
41
42 471 2011; Kos & Kramberger, 2017; Srivastava et al., 2015; Whiteside et al., 2017). A single sensor
43
44 472 may also be mounted in a low-profile manner on sporting equipment (Groh et al., 2016, 2017,
45
46 473 2015; Jensen et al., 2015). Unobtrusive use of a single IMU to capture generalised movements
47
48 474 across the whole body was demonstrated, with an IMU mounted on the posterior head in
49
50 475 swimming (Jensen et al., 2016, 2013), lower back during running (Kobsar et al., 2014), and
51
52 476 between the shoulder blades in rugby union (Kelly et al., 2012).

53
54
55
56
57 477 The majority of vision-based studies opted for a single camera set-up of RGB modality.
58
59 478 Data output from a single camera as opposed to multiple minimises the volume of data to process,
60
61 479 therefore reducing computational effort. However, detailed features may go uncaptured,
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

480 particularly in team sport competition which consists of multiple individuals participating in the
481 capture space at one time. In contrast, a multiple camera set-up reduces limitations including
482 occlusion and viewpoint variations. However, this may also increase the complexity of the
483 processing and model computational stages. Therefore, a trade-off between computational demands
484 and movement recording accuracy often needs to be made. As stated earlier, the placement of
485 cameras needs to suit the biomechanical nature of the targeted movement and the environment
486 situated in. Common camera capture systems used in sports science research such as Vicon Nexus
487 (Oxford, UK) and OptiTrack (Oregon, USA) were not present in this review. As this review
488 targeted studies investigating during on-field or in-situation sporting contexts, efficiency in data
489 collection is key for routine applications in training and competition. A simple portable RGB
490 camera is easy to set-up in a dynamic and changing environment, such as different soccer pitches,
491 rather than a multiple capture system such as Vicon that requires calibrated precision and are
492 substantially more expensive.

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

505 Supervised learning approaches were dominant across IMU and vision-based studies. This
506 is a method which involves a labelled ground truth training dataset typically manually annotated by
507 sport analysts. Labelled data instances were recorded as up to 15, 000 for vision-based (Victor et
508 al., 2017) and 416, 737 for sensor-based (Rassem et al., 2017) studies. Generation of a training data
509 set for supervised learning can be a tedious and labour-intensive task. It is further complicated if

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

510 multiple sensors or cameras are incorporated for several targeted movements. A semi-supervised or
511 unsupervised learning approach may be advantageous as data labelling is minimal or not required,
512 potentially reducing human errors in annotation. An unsupervised approach could suit specific
513 problems to explain key data features, via clustering (Mohammed et al., 2016; Sze et al., 2017).
514 Results computed by an unsupervised model (Kos, Ženko, Vlaj, & Kramberger, 2016) for tennis
515 serve, forehand and backhand stroke classification compared favourably well against a proposed
516 supervised approach (Connaghan et al., 2011).

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

540 (Ravi et al., 2016; J. B. Yang et al., 2015; Zebin et al., 2016; Zeng et al., 2014; Zheng, Liu, Chen,
541 Ge, & Zhao, 2014) and computer vision (Ji et al., 2013; Krizhevsky et al., 2012; LeCun et al.,
542 2015). As machine learning algorithms extract heuristic features requiring domain knowledge, this
543 creates shallower features which can make it harder to infer high-level and context aware activities
544 (J. B. Yang et al., 2015). Given the previously described advantages of deep learning algorithms
545 which apply to CNN, and the recent results of deep learning, future model developments may
546 benefit from exploring these methods in comparison to current bench mark models.

547 Model performance outcome metrics quantify and visualise the error rate between the
548 predicted outcome and true measure. Comparatively, a kernel form of an SVM was the most
549 common classifier implemented and produced the strongest machine learning approach model
550 prediction accuracies across both IMU (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Buthe
551 et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Pernek et al., 2015; Salman et al.,
552 2017; Schuldhaus et al., 2015; Whiteside et al., 2017) and vision-based study designs (Horton et
553 al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Reily et al., 2017; Shah et
554 al., 2007; Zhu et al., 2006). Classification accuracy was the most common reported measure
555 followed by confusion matrices, as ways to clearly present prediction results and derive further
556 measures of performance. Further measures included sensitivity (also called recall), specificity and
557 precision, whereby results closer to 1.0 indicate superior model performance, compared to 0.0 or
558 poor model performance. The F1-score (also called a F-measure or F-score) conveys the balances
559 between the precision and sensitivity of a model. An in-depth analysis performance metrics
560 specific to human activity recognition is located elsewhere (Minnen, Westeyn, Starner, Ward, &
561 Lukowicz, 2006; Ward, Lukowicz, & Gellersen, 2011). Use of specific evaluation methods
562 depends upon the data type. Conventional performance measures of error rate are generally
563 unsuitable for models developed from skewed training data (Provost & Fawcett, 2001). Using
564 conventional performance measures in this context will only take the default decision threshold on
565 a model trained, if there is an uneven class distribution this may lead to imprecision (Provost &
566 Fawcett, 2001; Seiffert, Khoshgoftaar, Van Hulse, & Napolitano, 2008). Alternative evaluators
567 including Receiver Operating Characteristics (ROC) curves and its single numeric measure, Area
568 Under ROC Curve (AUC), report model performances across all decision thresholds (Seiffert et al.,
569 2008). Making evaluations between study methodology have inherent complications due to each

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

570 formulating their own experimental parameter settings, feature vectors and training algorithms for
571 movement recognition. The No-Free-Lunch theorems are important deductions in the formation of
572 models for supervised machine learning (David H. Wolpert, 1996), and search and optimisation
573 algorithms (D H Wolpert & Macready, 1997). The theorems broadly reference that there is no ‘one
574 model’ that will perform optimally across all recognition problems. Therefore, experiments with
575 multiple model development methods for a particular problem is recommended. The use of prior
576 knowledge about the task should be implemented to adapt the model input and model parameters in
577 order to improve overall model success (Shalev-Shwartz & Ben-David, 2014).

578 Acquisition of athlete specific information, including statistics on number, type and
579 intensity of actions, may be of use in the monitoring of athlete load. Other potential applications
580 include personalised movement technique analysis (M. O’Reilly et al., 2017), automated
581 performance evaluation scoring (Reily et al., 2017) and team ball sports pass quality rating (Horton
582 et al., 2014). However, one challenge lies in delivering consistent, individualised models across
583 team field sports that are dynamic in nature. For example, classification of soccer shots and passes
584 showed a decline in model performance accuracy from a closed environment to a dynamic match
585 setting (Schuldhaus et al., 2015). A method to overcome accuracy limitations in dynamic team field
586 sports associated with solely using IMUs or vision may be to implement data fusion (Ó Conaire et
587 al., 2010). Furthermore, vision and deep learning approaches have demonstrated the ability to track
588 and classify team sport collective court activities and individual player specific movements in
589 volleyball (Ibrahim et al., 2016), basketball (Ramanathan et al., 2015) and ice hockey (Tora et al.,
590 2017). Accounting for methods from experimental set-up to model evaluation, previous reported
591 models should be considered and adapted based on the current problem. Furthermore, the balance
592 between model computational efficiency, results accuracy and complexity trade-offs calculations
593 are an important factor.

594 In the present study, meta-analysis was considered however variability across developed
595 model parameter reporting and evaluation methods did not allow for this to be undertaken. As this
596 field expands and further methodological approaches are investigated, it would be practical to
597 review analysis approaches both within and between sports. This review was delimited to machine
598 and deep learning approaches to sport movement detection and recognition. However, statistical or
599 parametric approaches not considered here such as discriminative functional analysis may also

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

600 show efficacy for sport-specific movement recognition. However, as the field of machine learning
601 is a rapidly developing area shown to produce superior results, a review encompassing all possible
602 other methods may have complicated the reporting. Since sport-specific movements and their
603 environments alter the data acquisition and analysis, the sports and movements reported in the
604 present study provide an overview of the current field implementations.

606 **5 Conclusions**

607
608 This systematic review reported on the literature using machine and deep learning methods to
609 automate sport-specific movement recognition. In addressing the research questions, both IMUs
610 and computer vision have demonstrated capacity in improving the information gained from sport
611 movement and skill recognition for performance analysis. A range of methods for model
612 development were used across the reviewed studies producing varying results. Conventional
613 machine learning algorithms such as Support Vector Machines and Neural Networks were most
614 commonly implemented. Yet in those studies which applied deep learning algorithms such as
615 Convolutional Neural Networks, these methods outperformed the machine learning algorithms in
616 comparison. Typically, the models were evaluated using a leave-one-out cross validation method
617 and reported model performances as a classification accuracy score. Intuitively, the adaptation of
618 experimental set-up, data processing, and recognition methods used are best considered in relation
619 to the characteristics of the sport and targeted movement(s). Consulting current models within or
620 similar to the targeted sport and movement is of benefit to address bench mark model performances
621 and identify areas for improvement. The application within the sporting domain of machine
622 learning and automated sport analysis coding for consistent uniform usage appears currently a
623 challenging prospect, considering the dynamic nature, equipment restrictions and varying
624 environments arising in different sports.

625 Future work may look to adopt, adapt and expand on current models associated with a
626 specific sports movement to work towards flexible models for mainstream analysis
627 implementation. Investigation of deep learning methods in comparison to conventional machine
628 learning algorithms would be of particular interest to evaluate if the trend of superior performances
629 is beneficial for sport-specific movement recognition. Analysis as to whether IMUs and vision

630 alone or together yield enhanced results in relation to a specific sport and its implementation
631 efficiency would also be of value. In consideration of the reported study information, this review
632 can assist future researchers in broadening investigative approaches for sports performance analysis
633 as a potential to enhancing upon current methods.

634

635 **Acknowledgements**

636 Not applicable.

637

638 **Disclosure statement**

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

641

642 **Funding**

643 The authors received no specific funding for this work.

644

645 **Author ORCID**

646 Sam Robertson 0000-0002-8330-0011

647 Alice Sweeting 0000-0002-9185-6773

648 Emily Cust 0000-0001-6927-6329

649

650 **References**

651

- 652 Adelsberger, R., & Tröster, G. (2013). Experts lift differently: Classification of weight-lifting
653 athletes. In *2013 IEEE International Conference on Body Sensor Networks* (pp. 1–6).
654 Cambridge, MA: Body Sensor Networks (BSN). <https://doi.org/10.1109/BSN.2013.6575458>
- 655 Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern*
656 *Recognition Letters*, 48, 70–80. <https://doi.org/10.1016/j.patrec.2014.04.011>
- 657 Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2017). Wearable motion
658 sensor based analysis of swing sports. In *2017 16th IEEE International Conference on*
659 *Machine Learning and Applications (ICMLA)* (pp. 261–267).
660 <https://doi.org/10.1109/ICMLA.2017.0-149>
- 661 Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports*
662 *Medicine*, 38(12), 1025–1043. <https://doi.org/10.2165/00007256-200838120-00006>
- 663 Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer*
664 *Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
665 *Bioinformatics)*, 7978 LNAI, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1
- 666 Bertasius, G., Park, H. S., Yu, S. X., & Shi, J. (2017). Am I a baller? Basketball performance
667 assessment from first-person videos. *Proceedings of the IEEE International Conference on*
668 *Computer Vision*, 2196–2204. <https://doi.org/10.1109/ICCV.2017.239>
- 669 Brock, H., & Ohgi, Y. (2017). Assessing motion style errors in ski jumping using inertial sensor
670 devices. *IEEE Sensors Journal*, (99), 1–11. <https://doi.org/10.1109/JSEN.2017.2699162>
- 671 Brock, H., Ohgi, Y., & Lee, J. (2017). Learning to judge like a human: convolutional networks for
672 classification of ski jumping errors. *Proceedings of the 2017 ACM International Symposium*

- 673 on *Wearable Computers - ISWC '17*, 106–113. <https://doi.org/10.1145/3123021.3123038>
- 674 Buckley, C., O'Reilly, M. A., Whelan, D., Vallely Farrell, A., Clark, L., Longo, V., ... Caulfield,
675 B. (2017). Binary classification of running fatigue using a single inertial measurement unit. In
676 *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor*
677 *Networks* (pp. 197–201). IEEE. <https://doi.org/10.1109/BSN.2017.7936040>
- 678 Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-
679 worn inertial sensors. *ACM Computing Surveys*, 46(3), 1–33.
680 <https://doi.org/http://dx.doi.org/10.1145/2499621>
- 681 Buthe, L., Blanke, U., Capkevics, H., & Tröster, G. (2016). A wearable sensing system for timing
682 analysis in tennis. In *BSN 2016 - 13th Annual Body Sensor Networks Conference* (pp. 43–48).
683 San Francisco, CA. <https://doi.org/10.1109/BSN.2016.7516230>
- 684 Bux, A., Angelov, P., & Habib, Z. (2017). Vision based human activity recognition: A review. In
685 P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in Computational Intelligence*
686 *Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence*
687 (pp. 341–371). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-](https://doi.org/10.1007/978-3-319-46562-3_23)
688 [46562-3_23](https://doi.org/10.1007/978-3-319-46562-3_23)
- 689 Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field
690 use of wearable inertial sensors for sport performance evaluation: a systematic review.
691 *Sensors*, 18(3), 873. <https://doi.org/10.3390/s18030873>
- 692 Chambers, R., Gabbett, T., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to
693 quantify sport-specific movements. *Sports Medicine*, 45(7), 1065–1081.
694 <https://doi.org/10.1007/s40279-015-0332-9>
- 695 Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., Walsh, M., & O'Mathuna, C. (2011).
696 Multi-sensor classification of tennis strokes. *Journal of IEEE Sensors*, 1437–1440.
- 697 Couceiro, M. S., Dias, G., Mendes, R., & Araújo, D. (2013). Accuracy of pattern detection
698 methods in the performance of golf putting. *Journal of Motor Behavior*, 45(1), 37–53.
699 <https://doi.org/10.1080/00222895.2012.740100>
- 700 Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic
701 recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement*
702 *Science*, 34(1), 63–80. <https://doi.org/10.1016/j.humov.2014.01.001>
- 703 Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for
704 context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7),
705 645–662. <https://doi.org/10.1007/s00779-010-0293-9>
- 706 Fong, D. T.-P., & Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower
707 limb biomechanics studies: A systematic review. *Sensors*, 10(12), 11556–11565.
708 <https://doi.org/10.3390/s101211556>
- 709 Gabbett, T., Jenkins, D., & Abernethy, B. (2012). Physical demands of professional rugby league
710 training and competition using microtechnology. *Journal of Science and Medicine in Sport*,
711 15, 80–86. <https://doi.org/10.1016/j.jsams.2011.07.004>
- 712 Gabbett, T., Jenkins, D. G., & Abernethy, B. (2011). Physical collisions and injury in professional
713 rugby league match-play. *Journal of Science and Medicine in Sport*, 14, 210–215.
714 <https://doi.org/10.1016/j.jsams.2011.01.002>
- 715 Gustin, P. B., McLean, O. C., Breed, R. V., & Spittle, M. (2014). Tackle and impact detection in
716 elite Australian football using wearable microsensor technology. *Journal of Sports Sciences*,
717 32(10), 947–953. <https://doi.org/10.1080/02640414.2013.868920>
- 718 Gustin, P. B., McLean, O. C., Spittle, M., & Breed, R. V. (2013). Quantification of tackling
719 demands in professional Australian football using integrated wearable athlete tracking
720 technology. *Journal of Science and Medicine in Sport*, 16(6), 589–593.
721 <https://doi.org/10.1016/j.jsams.2013.01.007>
- 722 Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes
723 using principal component analysis. *Journal of Sports Sciences*, 36(2), 229–237.
724 <https://doi.org/10.1080/02640414.2017.1298826>
- 725 Groh, B. H., Fleckenstein, M., & Eskofier, B. M. (2016). Wearable trick classification in freestyle
726 snowboarding. In *13th International Conference on Wearable and Implantable Body Sensor*
727 *Networks (BSN)* (pp. 89–93). IEEE. <https://doi.org/10.1109/BSN.2016.7516238>
- 728 Groh, B. H., Fleckenstein, M., Kautz, T., & Eskofier, B. M. (2017). Classification and visualization
729 of skateboard tricks using wearable sensors. *Pervasive and Mobile Computing*, 40, 42–55.
730 <https://doi.org/10.1016/j.pmcj.2017.05.007>
- 731 Groh, B. H., Kautz, T., & Schuldhuis, D. (2015). IMU-based trick classification in skateboarding.

- 732 In *KDD Workshop on Large-Scale Sports Analytics*.
- 733 Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods
734 to Oyama karate techniques recognition. *Symmetry*, 7(4), 1670–1698.
735 <https://doi.org/10.3390/sym7041670>
- 736 Hafer, J. F., & Boyer, K. A. (2017). Variability of segment coordination using a vector coding
737 technique: reliability analysis for treadmill walking and running. *Gait and Posture*, 51, 222–
738 227. <https://doi.org/10.1016/j.gaitpost.2016.11.004>
- 739 Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Proceedings Of The
740 International Joint Conference On Neural Networks, 1*, 593–605.
741 <https://doi.org/10.1109/IJCNN.1989.118638>
- 742 Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, 9(8),
743 1–32. <https://doi.org/10.1162/neco.1997.9.8.1735>
- 744 Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2014). Classification of passes in
745 football matches using spatiotemporal data. *ArXiv Preprint ArXiv:1407.5093*.
746 <https://doi.org/10.1145/3105576>
- 747 Howe, S. T., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cavanagh, B. P. (2017).
748 Quantifying important differences in athlete movement during collision-based team sports:
749 Accelerometers outperform global positioning systems. In *2017 IEEE International
750 Symposium on Inertial Sensors and Systems* (pp. 1–4). Kauai, HI, USA: IEEE.
751 <https://doi.org/10.1109/ISISS.2017.7935655>
- 752 Hulin, B. T., Gabbett, T., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can
753 accurately identify collision events during professional rugby league match-play. *Journal of
754 Science and Medicine in Sport*, 20(7), 638–642.
755 <https://doi.org/http://dx.doi.org/10.1016/j.jsams.2016.11.006>
- 756 Ibrahim, M., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A Hierarchical Deep
757 Temporal Model for Group Activity Recognition. *Cvpr*, 1971–1980.
758 <https://doi.org/10.1109/CVPR.2016.217>
- 759 Jensen, U., Blank, P., Kugler, P., & Eskofier, B. M. (2016). Unobtrusive and energy-efficient
760 swimming exercise tracking using on-node processing. *IEEE Sensors Journal*, 16(10), 3972–
761 3980. <https://doi.org/10.1109/JSEN.2016.2530019>
- 762 Jensen, U., Prade, F., & Eskofier, B. M. (2013). Classification of kinematic swimming data with
763 emphasis on resource consumption. In *2013 IEEE International Conference on Body Sensor
764 Networks, BSN 2013*. <https://doi.org/10.1109/BSN.2013.6575501>
- 765 Jensen, U., Schmidt, M., Hennig, M., Dassler, F. A., Jaitner, T., & Eskofier, B. M. (2015). An
766 IMU-based mobile system for golf putt analysis. *Sports Engineering*, 18(2), 123–133.
767 <https://doi.org/10.1007/s12283-015-0171-9>
- 768 Ji, S., Yang, M., Yu, K., & Xu, W. (2013). 3D convolutional neural networks for human action
769 recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 221–
770 231. <https://doi.org/10.1109/TPAMI.2012.59>
- 771 Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Multi-sensor Golf Swing Classification
772 Using Deep CNN. *Procedia Computer Science*, 129, 59–65.
773 <https://doi.org/10.1016/j.procs.2018.03.046>
- 774 Kapela, R., Świetlicka, A., Rybarczyk, A., Kolanowski, K., & O'Connor, N. E. (2015). Real-time
775 event classification in field sport videos. *Signal Processing: Image Communication*, 35, 35–
776 45. <https://doi.org/10.1016/j.image.2015.04.005>
- 777 Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014a). Large-
778 scale video classification with convolutional neural networks. *Computer Vision and Pattern
779 Recognition (CVPR), 2014 IEEE Conference On*, 1725–1732.
780 <https://doi.org/10.1109/CVPR.2014.223>
- 781 Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014b). Large-
782 scale video classification with convolutional nural networks. Retrieved December 18, 2017,
783 from <http://cs.stanford.edu/people/karpathy/deepvideo/>
- 784 Kasiri-Bidhendi, S., Fookes, C., Morgan, S., Martin, D. T., & Sridharan, S. (2015). Combat sports
785 analytics: Boxing punch classification using overhead depth imagery. In *2015 IEEE
786 International Conference on Image Processing (ICIP)* (pp. 4545–4549). Quebec City,
787 Canada: IEEE. <https://doi.org/10.1109/ICIP.2015.7351667>
- 788 Kasiri, S., Fookes, C., Sridharan, S., & Morgan, S. (2017). Fine-grained action recognition of
789 boxing punches from depth imagery. *Computer Vision and Image Understanding*, 159, 143–
790 153. <https://doi.org/10.1016/j.cviu.2017.04.007>

- 791 Kautz, T. (2017). Acquisition, filtering and analysis of positional and inertial data in sports. *FAU*
792 *Studies in Computer Science*, 2.
- 793 Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity
794 recognition in beach volleyball using a deep convolutional neural network. *Data Mining and*
795 *Knowledge Discovery*, 1–28. <https://doi.org/10.1007/s10618-017-0495-0>
- 796 Ke, S. R., Thuc, H., Lee, Y. J., Hwang, J. N., Yoo, J. H., & Choi, K. H. (2013). A review on video-
797 based human activity recognition. *Computers*, 2, 88–131.
798 <https://doi.org/10.3390/computers2020088>
- 799 Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions
800 in elite level rugby union using a wearable sensing device. *Sports Engineering*, 15(2), 81–92.
801 Retrieved from [https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-](https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5)
802 [0088-5](https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5)
- 803 Kobsar, D., Osis, S. T., Hettinga, B. A., & Ferber, R. (2014). Classification accuracy of a single tri-
804 axial accelerometer for training background and experience level in runners. *Journal of*
805 *Biomechanics*, 47(10), 2508–2511. <https://doi.org/10.1016/j.jbiomech.2014.04.017>
- 806 Kos, M., & Kramberger, I. (2017). A wearable device and system for movement and biometric data
807 Acquisition for sports applications. *IEEE Access*, 1–1.
808 <https://doi.org/10.1109/ACCESS.2017.2675538>
- 809 Kos, M., Ženko, J., Vlaj, D., & Kramberger, I. (2016). Tennis stroke detection and classification
810 using miniature wearable IMU device. In *International Conference on Systems, Signals, and*
811 *Image Processing*. <https://doi.org/10.1109/IWSSIP.2016.7502764>
- 812 Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of
813 classification techniques. *Informatica*, 31, 501–520. <https://doi.org/10.1115/1.1559160>
- 814 Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep
815 convolutional neural networks. *Advances In Neural Information Processing Systems*, 1097–
816 1105. <https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007>
- 817 Lai, D. T. H., Hetchl, M., Wei, X., Ball, K., & McLaughlin, P. (2011). On the difference in swing
818 arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors.
819 *Procedia Engineering*, 13, 219–225. <https://doi.org/10.1016/j.proeng.2011.05.076>
- 820 LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to
821 document recognition. *IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- 822 LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998). Efficient backprop. In *Neural Networks:*
823 *Tricks of the Trade* (Vol. 1524, pp. 9–50).
- 824 LeCun, Y., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, 521(7553), 436–444.
825 <https://doi.org/10.1038/nature14539>
- 826 Li, J., Tian, Q., Zhang, G., Zheng, F., Lv, C., & Wang, J. (2018). Research on hybrid information
827 recognition algorithm and quality of golf swing. *Computers and Electrical Engineering*, 1–
828 13. <https://doi.org/10.1016/j.compeleceng.2018.02.013>
- 829 Liao, W. H., Liao, Z. X., & Liu, M. J. (2003). Swimming style classification from video sequences.
830 In Kinmen (Ed.), *16th IPPR Conference on Computer Vision, Graphics and Image*
831 *Processing* (pp. 226–233). ROC.
- 832 Lu, W. L., Okuma, K., & Little, J. J. (2009). Tracking and recognizing actions of multiple hockey
833 players using the boosted particle filter. *Image and Vision Computing*, 27(1–2), 189–205.
834 <https://doi.org/10.1016/j.imavis.2008.02.008>
- 835 Magalhaes, F. A. de, Vannozzi, G., Gatta, G., & Fantozzi, S. (2015). Wearable inertial sensors in
836 swimming motion analysis: A systematic review. *Journal of Sports Sciences*, 33(7), 732–745.
837 <https://doi.org/10.1080/02640414.2014.962574>
- 838 Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical
839 activity from on-body accelerometers. *Sensors*, 10(2), 1154–1175.
840 <https://doi.org/10.3390/s100201154>
- 841 McNamara, D. J., Gabbett, T., Blanch, P., & Kelly, L. (2017). The relationship between wearable
842 microtechnology device variables and cricket fast bowling intensity. *International Journal of*
843 *Sports Physiology and Performance*, 1–20. [https://doi.org/https://doi.org/10.1123/ijsp.2016-](https://doi.org/https://doi.org/10.1123/ijsp.2016-0540)
844 [0540](https://doi.org/https://doi.org/10.1123/ijsp.2016-0540)
- 845 McNamara, D. J., Gabbett, T., Chapman, P., Naughton, G., & Farhart, P. (2015). The validity of
846 microsensors to automatically detect bowling events and counts in cricket fast bowlers.
847 *International Journal of Sports Physiology and Performance*, 10(1), 71–75.
848 <https://doi.org/10.1123/ijsp.2014-0062>
- 849 Minnen, D., Westeyn, T. L., Starner, T., Ward, J. a, & Lukowicz, P. (2006). Performance metrics

- 850 and evaluation issues for continuous activity recognition. In *Proc. Int. Workshop on*
851 *Performance Metrics for Intelligent Systems* (pp. 141–148).
852 <https://doi.org/10.1145/1889681.1889687>
- 1 853 Mitchell, E., Monaghan, D., & O'Connor, N. E. (2013). Classification of sporting activities using
2 854 smartphone accelerometers. *Sensors (Basel, Switzerland)*, *13*(4), 5317–5337.
3 855 <https://doi.org/10.3390/s130405317>
- 4 856 Mohammed, M., Khan, M., & Bashier, E. (2016). *Machine Learning: Algorithms and Applications*.
5 857 Milton: CRC Press.
- 6 858 Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting
7 859 items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7),
8 860 e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- 9 861 Montoliu, R., Martín-Félez, R., Torres-Sospedra, O., & Martínez-Usó, A. (2015). Team activity
10 862 recognition in Association football using a bag-of-words-based method. *Human Movement*
11 863 *Science*, *41*, 165–178. <https://doi.org/10.1016/j.humov.2015.03.007>
- 12 864 Mooney, R., Corley, G., Godfrey, A., Quinlan, L. R., & ÓLaighin, G. (2015). Inertial sensor
13 865 technology for elite swimming performance analysis: A systematic review. *Sensors*, *16*(1),
14 866 18. <https://doi.org/10.3390/s16010018>
- 15 867 Nibali, A., He, Z., Morgan, S., & Greenwood, D. (2017). Extraction and classification of diving
16 868 clips from continuous video footage. *ArXiv, pre-print*. Retrieved from
17 869 <https://arxiv.org/pdf/1705.09003.pdf>
- 18 870 O'Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017a). Classification
19 871 of lunge biomechanics with multiple and individual inertial measurement units. *Sports*
20 872 *Biomechanics*, *16*(3), 342–360. <https://doi.org/10.1080/14763141.2017.1314544>
- 21 873 O'Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017b). Technology in
22 874 strength and conditioning tracking lower-limb exercises with wearable sensors. *Journal of*
23 875 *Strength and Conditioning Research*, *31*(6), 1726–1736.
- 24 876 O'Reilly, M., Caulfield, B., Ward, T., Johnston, W., & Doherty, C. (2018). Wearable Inertial
25 877 Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review.
26 878 *Sports Medicine*. <https://doi.org/10.1007/s40279-018-0878-4>
- 27 879 O'Reilly, M., Whelan, D., Chaniyalidis, C., Friel, N., Delahunt, E., Ward, T., & Caulfield, B.
28 880 (2015). Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE*
29 881 *12th International Conference on Wearable and Implantable Body Sensor Networks*. IEEE.
30 882 <https://doi.org/10.1109/BSN.2015.7299380>
- 31 883 O'Reilly, M., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017). Classification of
32 884 deadlift biomechanics with wearable inertial measurement units. *Journal of Biomechanics*,
33 885 *58*, 155–161. <https://doi.org/10.1080/14763141.2017.1314544>
- 34 886 Ó Conaire, C., Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., & Buckley, J. (2010).
35 887 Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings*
36 888 *of the first ACM international workshop on Analysis and retrieval of tracked events and*
37 889 *motion in imagery streams* (pp. 51–56). ACM. <https://doi.org/10.1145/1877868.1877882>
- 38 890 Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength
39 891 training exercises with wearable sensors. *Journal of Biomedical Informatics*, *58*, 145–155.
40 892 <https://doi.org/10.1016/j.jbi.2015.09.020>
- 41 893 Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in
42 894 ubiquitous computing. *International Joint Conference on Artificial Intelligence (IJCAI)*,
43 895 1729.
- 44 896 Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision*
45 897 *Computing*, *28*(6), 976–990. <https://doi.org/10.1016/j.imavis.2009.11.014>
- 46 898 Preece, S. J., Goulermas, J. Y., Kenney, L., & Howard, D. (2009). A comparison of feature
47 899 extraction methods for the classification of dynamic activities from accelerometer data. *IEEE*
50 900 *Transactions on Biomedical Engineering*, *56*(3), 871–879.
51 901 <https://doi.org/10.1109/TBME.2008.2006190>
- 52 902 Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009).
53 903 Activity identification using body-mounted sensors: A review of classification techniques.
54 904 *Physiological Measurement*, *30*(4), R1–R33. <https://doi.org/10.1088/0967-3334/30/4/R01>
- 55 905 Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine*
56 906 *Learning*, *42*(3), 203–231. <https://doi.org/10.1023/A:1007601015854>
- 57 907 Qaisar, S., Imtiaz, S., Glazier, P., Farooq, F., Jamal, A., Iqbal, W., & Lee, S. (2013). A method for
58 908 cricket bowling action classification and analysis using a system of inertial sensors. In

- 909 *International Conference on Computational Science and its Applications* (pp. 396–412).
910 Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-39649-6>
- 911 Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2015).
912 Detecting events and key actors in multi-person videos.
913 <https://doi.org/10.1109/CVPR.2016.332>
- 914 Rassem, A., El-Beltagy, M., & Saleh, M. (2017). Cross-country skiing gears classification using
915 deep learning. *ArXiv Preprint ArXiv:1706.08924*. Retrieved from
916 <https://arxiv.org/pdf/1706.08924v1.pdf>
- 917 Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor
918 data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health*
919 *Informatics*, 21(1), 1–1. <https://doi.org/10.1109/JBHI.2016.2633287>
- 920 Reily, B., Zhang, H., & Hoff, W. (2017). Real-time gymnast detection and performance analysis
921 with a portable 3D camera. *Computer Vision and Image Understanding*, 159, 154–163.
922 <https://doi.org/10.1016/j.cviu.2016.11.006>
- 923 Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic
924 classification of sub-techniques in classical cross-country skiing using a machine learning
925 algorithm on micro-sensor data. *Sensors (Switzerland)*, 18(1), 75.
926 <https://doi.org/10.3390/s18010075>
- 927 Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using
928 deep learning neural networks. *Expert Systems with Applications*, 59, 235–244.
929 <https://doi.org/10.1016/j.eswa.2016.04.032>
- 930 Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events
931 in soccer videos. *Applied Artificial Intelligence*, 27(7), 656–667.
932 <https://doi.org/10.1080/08839514.2013.787779>
- 933 Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling
934 action in the game of cricket. *Data Mining and Knowledge Discovery*, 31(6), 1706–1734.
935 <https://doi.org/10.1007/s10618-017-0511-4>
- 936 Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B. M. (2015). Inertial
937 sensor-based approach for shot/ pass classification during a soccer match. In *Proc. 21st ACM*
938 *KDD Workshop on Large-Scale Sports Analytics* (pp. 1–4). Sydney, Australia.
- 939 Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving
940 classification performance when training data is skewed. In *9th International Conference on*
941 *Pattern Recognition* (pp. 1–4). <https://doi.org/10.1109/ICPR.2008.4761297>
- 942 Shah, H., Chokalingam, P., Paluri, B., & Pradeep, N. (2007). Automated stroke classification in
943 tennis. *Image Analysis and Recognition*, 1128–1137.
- 944 Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to*
945 *algorithms*. New York, USA: Cambridge University Press.
- 946 Sharma, M., Srivastava, R., Anand, A., Prakash, D., & Kaligounder, L. (2017). Wearable motion
947 sensor based phasic analysis of tennis serve for performance feedback. In *2017 IEEE*
948 *International Conference on Acoustics, Speech and Signal Processing* (pp. 5945–5949). New
949 Orleans, LA: IEEE.
- 950 Sprager, S., & Juric, M. B. (2015). *Inertial sensor-based gait recognition: A review. Sensors*
951 *(Switzerland)* (Vol. 15). <https://doi.org/10.3390/s150922089>
- 952 Srivastava, R., Patwari, A., Kumar, S., Mishra, G., Kaligounder, L., & Sinha, P. (2015). Efficient
953 characterization of tennis shots and game analysis using wearable sensors data. In *2015 IEEE*
954 *Sensors- Proceedings* (pp. 1–4). Busan. <https://doi.org/10.1109/ICSENS.2015.7370311>
- 955 Stein, M., Janetzko, H., Lamprecht, A., Breitreutz, T., Zimmermann, P., Goldlücke, B., ... Keim,
956 D. A. (2018). Bring it to the pitch: combining video and movement data to enhance team
957 sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 13–22.
958 <https://doi.org/10.1109/TVCG.2017.2745181>
- 959 Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks:
960 A tutorial and survey. *IEEE*, 105(2), 2295–2329. Retrieved from
961 <http://arxiv.org/abs/1703.09039>
- 962 Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports:
963 Current applications and research topics. *Computer Vision and Image Understanding*, 159, 3–
964 18. <https://doi.org/10.1016/j.cviu.2017.04.011>
- 965 Titterton, D. H., & Weston, J. L. (2009). *Strapdown inertial navigation technology* (2nd ed.).
966 Reston, VA: AIAA.
- 967 Tora, M. R., Chen, J., & Little, J. J. (2017). Classification of puck possession events in ice hockey.

- 968 In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*
969 *Workshops* (pp. 147–154). <https://doi.org/10.1109/CVPRW.2017.24>
- 970 Victor, B., He, Z., Morgan, S., & Miniutti, D. (2017). Continuous video to simple signals for
971 swimming stroke detection with convolutional neural networks. *ArXiv Preprint*
972 *ArXiv:1705.09894*. <https://doi.org/10.1111/j.1467-8330.1974.tb00606.x>
- 973 Wagner, D., Kalischewski, K., Velten, J., & Kummert, A. (2017). Activity recognition using
974 inertial sensors and a 2-D convolutional neural network. In IEEE (Ed.), *2017 10th*
975 *International Workshop on Multidimensional (nD) Systems (nDS)* (pp. 1–6).
976 <https://doi.org/10.1109/NDS.2017.8070615>
- 977 Wagner, J. F. (2018). About motion measurement in sports based on gyroscopes and
978 accelerometers - an engineering point of view. *Gyroscopy and Navigation*, *9*(1), 1–18.
979 <https://doi.org/10.1134/S2075108718010091>
- 980 Ward, J. A., Lukowicz, P., & Gellersen, H.-W. (2011). Performance metrics for activity
981 recognition. In *ACM Trans. on Intelligent Systems and Technology* (Vol. 2, pp. 111–132).
- 982 Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring hitting load in tennis using
983 inertial sensors and machine learning. *International Journal of Sports Physiology and*
984 *Performance*, 1–20. <https://doi.org/https://doi.org/10.1123/ijsp.2016-0683>
- 985 Wixted, A., Billing, D. C., & James, D. A. (2010). Validation of trunk mounted inertial sensors for
986 analysing running biomechanics under field conditions, using synchronously collected foot
987 contact data. *Sports Engineering*, *12*(4), 207–212. <https://doi.org/10.1007/s12283-010-0043-2>
- 988 Wixted, A., Portus, M., Spratford, W., & James, D. A. (2011). Detection of throwing in cricket
989 using wearable sensors. *Sports Technology*, *4*(3–4), 134–140.
990 <https://doi.org/10.1080/19346182.2012.725409>
- 991 Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural*
992 *Computation*, *8*(7), 1341–1390. <https://doi.org/10.1162/neco.1996.8.7.1391>
- 993 Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimisation. *IEEE*
994 *Transactions on Evolutionary Computation*, *1*(1), 67–82.
995 <https://doi.org/10.1023/A:1021251113462>
- 996 Wundersitz, D. W., Gastin, P. B., Richter, C., Robertson, S., & Netto, K. J. (2015). Validity of a
997 trunk-mounted accelerometer to assess peak accelerations during walking, jogging and
998 running. *European Journal of Sport Science*, *15*(5), 382–390.
999 <https://doi.org/10.1080/17461391.2014.955131>
- 1000 Wundersitz, D. W., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of
1001 a trunk-mounted accelerometer to measure peak impacts during team sport movements.
1002 *International Journal of Sports Medicine*, *36*(9), 742–746. <https://doi.org/10.1055/s-0035-1547265>
- 1003
- 1004 Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015).
1005 Classification of team sport activities using a single wearable tracking device. *Journal of*
1006 *Biomechanics*, *48*(15), 3975–3981. <https://doi.org/10.1016/j.jbiomech.2015.09.015>
- 1007 Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for
1008 physical activity monitoring. *Sensors*, *10*(8), 7772–7788. <https://doi.org/10.3390/s100807772>
- 1009 Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Shonali, K. (2015). Deep convolutional neural
1010 networks on multichannel time series for human activity recognition. In *Proceedings of the*
1011 *24th International Conference on Artificial Intelligence* (pp. 3995–4001).
- 1012 Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object
1013 interaction activities. In *Computer Vision and Pattern Recognition* (pp. 17–24). IEEE.
- 1014 Young, C., & Reinkensmeyer, D. J. (2014). Judging complex movement performances for
1015 excellence: a principal components analysis-based technique applied to competitive diving.
1016 *Human Movement Science*, *36*, 107–122. <https://doi.org/10.1016/j.humov.2014.05.009>
- 1017 Yu, G., Jang, Y. J., Kim, J., Kim, J. H., Kim, H. Y., Kim, K., & Panday, S. B. (2016). Potential of
1018 IMU sensors in performance analysis of professional alpine skiers. *Sensors (Switzerland)*,
1019 *16*(4), 1–21. <https://doi.org/10.3390/s16040463>
- 1020 Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human Activity Recognition with Inertial
1021 Sensors Using a Deep Learning Approach. *Proc. of IEEE Sensors 2016*, (1), 1–3.
1022 <https://doi.org/10.1109/ICSENS.2016.7808590>
- 1023 Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014).
1024 Convolutional neural networks for human activity recognition using mobile sensors. In
1025 *Proceedings of the 6th International Conference on Mobile Computing, Applications and*
1026 *Services* (pp. 197–205). <https://doi.org/10.4108/icst.mobibase.2014.257786>

1027 Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity
1028 recognition using vision-based method. *Journal of Healthcare Engineering*, 2017, 1–31.
1029 <https://doi.org/10.1155/2017/3090343>
1 1030 Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-
2 1031 channels deep convolutional neural networks. In *International Conference on Web-Age*
3 1032 *Information Management* (pp. 298–310). Springer. [https://doi.org/10.1007/978-3-319-08010-](https://doi.org/10.1007/978-3-319-08010-9_33)
4 1033 [9_33](https://doi.org/10.1007/978-3-319-08010-9_33)
5 1034 Zhu, G., Xu, C., Gao, W., & Huang, Q. (2006). Action recognition in broadcast tennis video.
6 1035 *Computer Vision in Human-Computer Interaction*, 89–98.
7 1036 https://doi.org/10.1007/11754336_9
8 1037 Ziaefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review.
9 1038 *Pattern Recognition*, 48(8), 2329–2345. <https://doi.org/10.1016/j.patcog.2015.03.006>
10 1039

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1 - PRISMA flow diagram

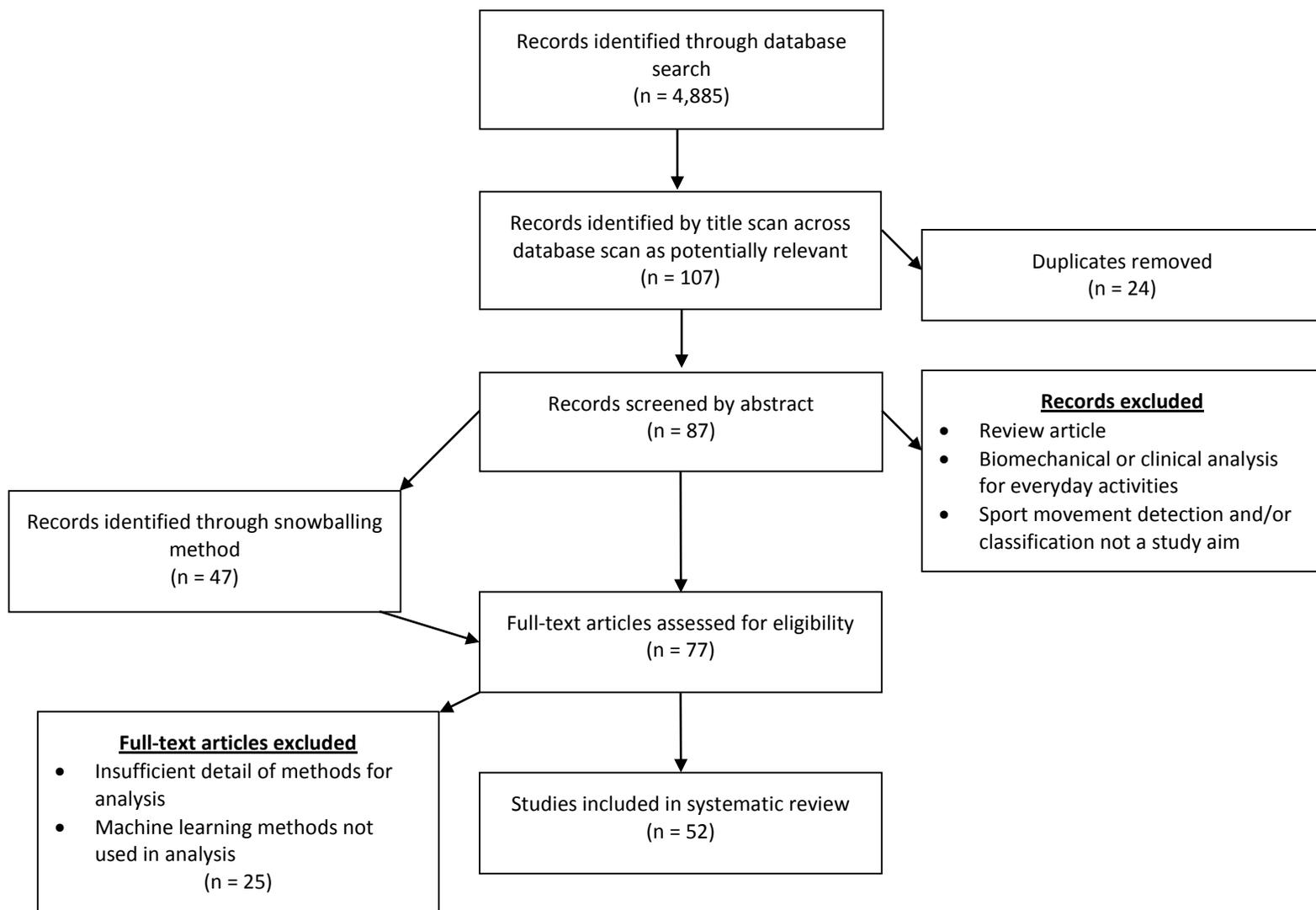


Figure 1 PRISMA flow diagram for study search, screen and selection process.

Table 1. Key word search term strings per database.

Database key word searches
<p>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)</p> <hr/> <p>(((sport OR athlete* OR player*)) AND (video OR vision)) AND movement classification</p>
<p>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)</p> <p>(((((((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</p>
<p>ScienceDirect: ((sport OR athlete* OR player*)) and ((inertial sensor OR accelerometer)</p> <p>((sport OR athlete* OR player*)) and TITLE-ABSTR-KEY((vision OR video OR camera) AND (detection OR classification)).</p>
<p>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)</p> <hr/> <p>(((sport OR athlete* OR player*)) AND (video OR vision)) AND movement classification</p>
<p>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)</p> <hr/> <p>(((sport OR athlete* OR player*)) AND (video OR vision)) AND movement classification</p>
<p>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)</p> <p>(((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)</p>

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

Table 2 Study inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • 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 μ T)	Sample rate
(Adelsberger & Tröster, 2013)	Ethos	3	Left ankle, wrist, lower back	3	± 6 g	NR	3	± 2000 $^{\circ}$ /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 $^{\circ}$ /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 $^{\circ}$ /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 $^{\circ}$ /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 $^{\circ}$ /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 $^{\circ}$ /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 μ T)	Sample rate
(Groh, Kautz, & Schuldhuis, 2015)	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 μ T	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 μ T	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 μ T	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 μ T)	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 $^{\circ}$ /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	5 th lumbar vertebra	3	± 16 g	51.2 Hz	3	± 500 $^{\circ}$ /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 $^{\circ}$ /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 $^{\circ}$ /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 μ T)	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	$\pm 16g$	1000 Hz	NR	NR	NR	NR	NR	NR
(Srivastava et al., 2015)	Samsung Gear S smart watch	1	Wrist of racquet arm	3	$\pm 8 g$	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	$\pm 16 g$	500 Hz	3	± 2000 °/s	500 Hz	3	± 1200 μ T	500 Hz

g G-forces, *Ga* gauss, *Hz* Hertz, *IMU* inertial measurement unit, μ *T* 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
<i>fps</i> frames per second, <i>Hz</i> hertz, <i>MPEG</i> Moving Picture Experts Group, <i>RGB</i> red green blue <i>NR</i> 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	Recogniti on 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 α_i , amplitude offset β_i , standard deviation σ_i , and mean value μ_i	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	Recogniti on 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	Recogniti on 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	<i>Part A</i> Four: male snowboarders, as two regular and two goofy stance directions. <i>Part B</i> 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	Recogniti on 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	Recogniti on 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	Recogniti on 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	Recogniti on 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	Recogniti on 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, 25 th and 75 th 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, 25 th percentile, 75 th 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	Recogniti on 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	Recogniti on 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 (radial 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, *ReLU*s 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: <ul style="list-style-type: none"> • Precision 0.95 • Recall 0.96 • F1- score 0.96 CA: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 100% (when athletes did not move between reps) Classification: <ul style="list-style-type: none"> • CA 94.117% (between expert and beginner level) Classification: <ul style="list-style-type: none"> • 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 $93 \pm 0.08\%$	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: <ul style="list-style-type: none"> • LIMU lumbar spine CA 75% • IMU right shank CA 70% • IMU left shank CA 67% Personalised classifier: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • Overall 76% • Side steps 96% • Shot steps 63% LOOCV: <ul style="list-style-type: none"> • Precision $0.49 \pm 0.04\%$ • Recall $0.49 \pm 0.22\%$ User-specific: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • Candidate strokes 85% • Non-candidate strokes 85% Classification accuracy: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • Sensitivity 94.2% • Specificity 99.9% Classification: <ul style="list-style-type: none"> • CA 97.8% (NB and SVM) Computation effort (lowest): <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> Recall 0.99 Precision 0.368 Trick category classification: <ul style="list-style-type: none"> Grind recall 0.966 Grind precision 0.885 Airs recall 0.974 Airs precision 0.910 Trick class CA: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> Precision 0.669 Recall 0.964 Classification: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 18.1 ms (NB with feature selection) Class prediction computational time: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> Right arm data CA 89.41% (kNN) Full-body data CA 93.44% (kNN) 		Data fusion of accelerometer and vision data improved CA: <ul style="list-style-type: none"> Vision back viewpoint with full body accelerometer 100% CA (kNN) Data fusion overcame viewpoint sensitivity <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> Sensitivity 64.41% Specificity 88.01% CA 80.45% Multi-label classification; <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> CA 90% Sensitivity 80% Specificity 92% Classify specific technique deviations Five lower limb IMU set-up: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> Global classifier 73% (RF) Personalized classifier 84% (RF) Natural technique deviations multi-class CA: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> CA range $84.2 \pm 11.3\%$ to $93.6 \pm 0.5\%$ Intensity error: <ul style="list-style-type: none"> range 1.2% to $6.6 \pm 2.5\%$ 	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) <ul style="list-style-type: none"> 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 $81 \pm 3.12\%$ (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): <ul style="list-style-type: none"> • Leg type 99.9% • Other events 96.7% • Pass or shot 88.6% Match conditions CA (SVM): <ul style="list-style-type: none"> • Shot 86.7% • Pass 81.7% 	Video footage manually labelled	
(Srivastava et al., 2015)	Detection accuracy, CA		Shot detection accuracy: <ul style="list-style-type: none"> • Professional 99.58% • Novice 98.96% • Total 99.41% Shot CA: <ul style="list-style-type: none"> • 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): <ul style="list-style-type: none"> • Condition one $97.43 \pm 0.24\%$ • Condition two $93.21 \pm 0.45\%$ 	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
<p>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.</p>					

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
(Bertasio, Park, Yu, & Shi, 2017)	Basketball: somebody 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 5 th 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 the 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 frames. 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
<p>2D two dimensional, <i>BLSTM</i> bidirectional LSTM, <i>CNN</i> convolutional neural network, <i>DOI</i> Depth of interest segmentation, <i>DT</i> decision tree, <i>ELU</i> Exponential Linear Units, <i>FFT</i> Fast Fourier Transform, <i>GDL</i> Gesture Description Language, <i>HMM</i> Hidden Markov Model, <i>HOG</i> Histogram of Oriented Gradients, <i>HSV</i> Hue-Saturation-Value-Colour-Histogram, <i>kNN</i> k-Nearest Neighbour, <i>LDA</i> linear discriminative analysis, <i>LR</i> logistic regression, <i>LS-SVM</i> least squares support vector machine, <i>MLP</i> multi-layer perceptron, <i>NB</i> Naïve Bayesian, <i>NN</i> neural network, <i>PCA</i> principal component analysis, <i>PNG</i> Portable Network Graphics, <i>QDA</i> quadratic discriminative analysis, <i>RBF</i> radial basis function, <i>RF</i> random forests, <i>RUSBoost</i> Random Under Sampling Boosting, <i>SAD3D</i> Silhouette Activity Descriptor in 3 Dimensions, <i>SPPCA</i> Switching Probabilistic Principal Component Analysis, <i>SVM</i> Support Vector Machine, <i>SVR</i> Support Vector Regression.</p>							

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 $93 \pm 7\%$ 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	2/3 rd of total data as training set, 1/3 rd 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. (Karpaty 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 $92.89 \pm 0.2\%$	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: <ul style="list-style-type: none"> Vision back viewpoint with full body accelerometer CA 100% (kNN) Data fusion overcame viewpoint sensitivity <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> • Left recall 84.08%, • Left precision 89.80% • Right recall 90.20%, • Right precision 84.66%. Tennis stroke classification using action clips: <ul style="list-style-type: none"> • Left recall 87.50%, • Left precision 90.74% • Right recall 89.80%, • Right precision 86.27% 		
<p><i>CA</i> classification accuracy, <i>CNN</i> convolutional neural network, <i>DE</i> detected events, <i>DTE</i> detected true events, <i>GMM</i> Gaussian mixture model, <i>HMM</i> Hidden Markov Model, <i>kNN</i> k-Nearest Neighbour, <i>LOO-CV</i> leave-one-out cross validation, <i>LOSO-CV</i> leave-one-subject-out cross validation, <i>LS-SVM</i> least squares support vector machine, <i>NE</i> number of events, <i>RF</i> random forests, <i>ROC</i> receiver operation characteristic curve, <i>SVM</i> Support Vector Machine.</p>					

- 1 **Machine and deep learning for sport-specific movement recognition: a systematic review of**
- 2 **model development and performance**
- 3
- 4 **Running title:**
- 5 Machine and deep learning for sport movement recognition review
- 6

7 **Abstract**

8

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

24

25

26 **Key Words:**

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

29 **1. Introduction**

30

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

46 Human activities detected by inertial sensing devices and computer vision are represented
47 as wave signal features corresponding to specific actions, which can be logged and extracted. Human
48 movement activities are considered hierarchically structured and can be broken down to basic
49 movements. Therefore, the context of signal use, intra-class variability, and inter-class similarity
50 between activities require consideration during experimental set-up and model development.
51 Wearable IMUs contain a combination of accelerometer, gyroscope, and magnetometer sensors
52 measuring along one to three axes. These sensors quantify acceleration, angular velocity, and the
53 direction and orientation of travel respectively (Gastin, McLean, Breed, & Spittle, 2014). These
54 sensors can capture repeated movement patterns during sport training and competitions (Camomilla,
55 Bergamini, Fantozzi, & Vannozzi, 2018; Chambers, Gabbett, Cole, & Beard, 2015; J. F. Wagner,
56 2018). Advantages include being wireless, lightweight and self-contained in operation. Inertial
57 measurement units have been utilised in quantifying physical output and tackling impacts in
58 Australian Rules football (Gastin et al., 2014; Gastin, McLean, Spittle, & Breed, 2013) and rugby

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

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

83 Machine learning algorithms learn from data input for automated model building and
84 perform tasks without being explicitly programmed. The algorithm goal is to output a response
85 function $h\sigma(\bar{x})$ that will predict a ground truth variable y from an input vector of variables \bar{x} . Models
86 are run for classification techniques to predict a target class (Kotsiantis, Zaharakis, & Pintelas, 2007),
87 or regression to predict discrete or continuous values. Models are aimed at finding an optimal set of
88 parameters σ to describe the response function, and then make predictions on unseen unlabelled data

89 input. Within these, model training approaches can generally run as supervised learning,
90 unsupervised learning or semi-supervised learning (Mohammed, Khan, & Bashier, 2016; Sze, Chen,
91 Yang, & Emer, 2017).

92 Processing raw data is limited for conventional machine learning algorithms, as they are
93 unable to effectively be trained on abstract and high-dimensional data that is inconsistent, contains
94 missing values or noisy artefacts (Bux et al., 2017; Kautz, 2017). Consequently, several pre-
95 processing stages are required to create a suitable data form for input into the classifier algorithm
96 (Figo, Diniz, Ferreira, & Cardoso, 2010). Filtering (Figo et al., 2010; Wundersitz, Gastin, Robertson,
97 Davey, & Netto, 2015), window capture durations (Mitchell, Monaghan, & O'Connor, 2013; Preece,
98 Goulermas, Kenney, & Howard, 2009; Wundersitz, Josman, et al., 2015), and signal frequency cut-
99 offs (Wundersitz, Gastin, Richter, Robertson, & Netto, 2015; Wundersitz, Gastin, Robertson, et al.,
100 2015) are common techniques applied prior to data prior to dynamic human movement recognition.
101 Well-established filters for processing motion signal data include the Kalman filter (Kautz, 2017;
102 Titterton & Weston, 2009; D. Wagner, Kalischewski, Velten, & Kummert, 2017) and a Fourier
103 transform filter (Preece, Goulermas, Kenney, Howard, et al., 2009) such as a fast Fourier transform
104 (Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015; Preece, Goulermas, Kenney, &
105 Howard, 2009). Near real-time processing benefits from reducing memory requirements,
106 computational demands, and essential bandwidth during whole model implementation. Signal
107 feature extraction and selection favours faster processing by reducing the signals to the critical
108 features that can discriminate the targeted activities (Bulling et al., 2014). Feature extraction involves
109 identifying the key features that help maximise classifier success, and removing features that have
110 minimal impact in the model (Mannini & Sabatini, 2010). Thus, feature selection involves
111 constructing data representations in subspaces with reduced dimensions. These identified variables
112 are represented in a compact feature variable (Mannini & Sabatini, 2010). Common methods include
113 principal component analysis (PCA) (Gløersen, Myklebust, Hallén, & Federolf, 2018; Young &
114 Reinkensmeyer, 2014), vector coding techniques (Hafer & Boyer, 2017) and empirical cumulative
115 distribution functions (ECDF) (Plötz, Hammerla, & Olivier, 2011). An ECDF approach has been
116 shown to be advantageous over PCA as it derives representations of raw input independent of the
117 absolute data ranges, whereas PCA is known to have reduced performance when the input data is not
118 properly normalised (Plötz et al., 2011). For further detailed information on the acquisition, filtering

119 and analysis of IMU data for sports application and vision-based human activity recognition, see
120 (Kautz, 2017) and (Bux et al., 2017), respectively.

121 Deep learning is a division of machine learning, characterised by deeper neural network
122 model architectures and are inspired by the biological neural networks of the human brain (Bengio,
123 2013; LeCun et al., 2015; Sze et al., 2017). The deeper hierarchical models create a profound
124 architecture of multiple hidden layers based on representative learning with several processing and
125 abstraction layers (Bux et al., 2017; J. B. Yang et al., 2015). These computational models allow data
126 input features to be automatically extracted from raw data and transformed to handle unstructured
127 data, including vision (LeCun et al., 2015; Ravi, Wong, Lo, & Yang, 2016). This direct input avoids
128 several processing steps required in machine learning during training and testing, therefore reducing
129 overall computational times. A current key element within deep learning is backpropagation (Hecht-
130 Nielsen, 1989; LeCun, Bottou, Orr, & Müller, 1998). Backpropagation is a fast and computationally
131 efficient algorithm, using gradient descent, that allows training deep neural networks to be tractable
132 (Sze et al., 2017). Human activity recognition has mainly been performed using conventional
133 machine learning classifiers. Recently, deep learning techniques have enhanced the bench mark and
134 applications for IMUs (Kautz et al., 2017; Ravi et al., 2016; Ronao & Cho, 2016; J. B. Yang et al.,
135 2015; Zebin, Scully, & Ozanyan, 2016; Zeng et al., 2014) and vision (Ji, Yang, Yu, & Xu, 2013;
136 Karpathy et al., 2014a; Krizhevsky, Sutskever, & Hinton, 2012; Nibali, He, Morgan, & Greenwood,
137 2017) in human movement recognition producing more superior model performance accuracy.

138 The objective of this study was to systematically review the literature investigating sport-
139 specific automated movement detection and recognition. The review focusses on the various
140 technologies, analysis techniques and performance outcome measures utilised. There are several
141 reviews within this field that are sensor-based including wearable IMUs for lower limb biomechanics
142 and exercises (Fong & Chan, 2010; M. O'Reilly, Caulfield, Ward, Johnston, & Doherty, 2018),
143 swimming analysis (Magalhaes, Vannozzi, Gatta, & Fantozzi, 2015; Mooney et al., 2015),
144 quantifying sporting movements (Chambers et al., 2015) and physical activity monitoring (C. C.
145 Yang & Hsu, 2010). A recent systematic review has provided an evaluation on the in-field use of
146 inertial-based sensors for various performance evaluation applications (Camomilla et al., 2018).
147 Vision-based methods for human activity recognition (Aggarwal & Xia, 2014; Bux et al., 2017; Ke
148 et al., 2013; Zhang et al., 2017), semantic human activity recognition (Ziaeefard & Bergevin, 2015)

149 and motion analysis in sport (Barris & Button, 2008) have also been reviewed. However, to date,
150 there is no systematic review across sport-specific movement detection and recognition via machine
151 or deep learning. Specifically, incorporating IMUs and vision-based data input, focussing on in-field
152 applications as opposed to laboratory-based protocols and detailing the analysis and machine
153 learning methods used.

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

161

162 **2. Methods**

163

164 **2.1 Search strategy**

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

175

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

177

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

179

180 **2.2 Data extraction**

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

208 attachment of sensors (IMUs), sample frequency, and sensor data types used in analysis (IMUs).
209 Studies identified and full data extracted were reviewed by a second author.

210

211 **3. Results**

212

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

218

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

220

221 **3.1 Experimental design**

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

238 (Srivastava et al., 2015). Semantic descriptors were reported for classification models that predicted
239 athlete training background, experience and fatigue level. These included running (Buckley et al.,
240 2017; Kobsar, Osis, Hettinga, & Ferber, 2014), rating of gymnastic routines (Reily, Zhang, & Hoff,
241 2017), soccer pass classification based on its quality (Horton, Gudmundsson, Chawla, & Estephan,
242 2014), cricket bowling legality (Qaisar et al., 2013; Salman, Qaisar, & Qamar, 2017), ski jump error
243 analysis (Brock & Ohgi, 2017; Brock, Ohgi, & Lee, 2017) and strength training technique deviations
244 (M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a; M. O'Reilly et al., 2015; M.
245 O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017). One approach (Yao & Fei-Fei, 2010),
246 encoded the mutual context of human pose and sporting equipment using semantics, to facilitate the
247 detection and classification of movements including a cricket bat and batsman coupled movements.

248 Total participant numbers for IMU-based studies ranged from one (Qaisar et al., 2013) to 30
249 (Kautz et al., 2017). Reported data individual instance sample sizes for sensor studies ranged from
250 150 (Salman et al., 2017) to 416, 737 (Rassem, El-Beltagy, & Saleh, 2017). Vision-based studies
251 that explicitly reported total participant details ranged from five (Ó Conaire et al., 2010) to 40 (Victor
252 et al., 2017). Vision dataset sample sizes varied across studies, from 50 individual action clips (Liao
253 et al., 2003) to 15, 000 (Victor et al., 2017). One study (Karpathy et al., 2014a) used the publicly
254 available Sports-1M, as previously described. Vision-based studies also reported datasets in total
255 time, 10.3 hours (Bertasius, Park, Yu, & Shi, 2017), 3 hours (Montoliu, Martín-Félez, Torres-
256 Sospedra, & Martínez-Usó, 2015), 1, 500 minutes (Shah et al., 2007), and 50 hours (Kapela et al.,
257 2015), and by frame numbers, 6, 035 frames (Zhu, Xu, Gao, & Huang, 2006) and 10, 115 frames
258 (Reily et al., 2017).

259

260 **3.2 Inertial measurement unit specifications**

261 A range of commercially available and custom-built IMUs were used in the IMU-based studies (n=
262 30), as presented in Table 3. Of these, 23% reported using a custom-built sensor. Of the IMU-based
263 studies, the number of sensors mounted or attached to each participant or sporting equipment piece
264 ranged from one to nine. The majority of studies (n= 22) provided adequate details of sensor
265 specifications including sensor type, axes, measurement range, and sample rate used. At least one
266 characteristic of sensor measurement range or sample rate used in data collection was missing from
267 eight studies. All studies used triaxial sensors and collected accelerometer data. For analysis and

268 model development, individual sensor data consisted of only accelerometer data (n = 8), both
269 accelerometer and gyroscope data (n = 15), and accelerometer, gyroscope and magnetometer data (n
270 = 7). The individual sensor measurement ranges reported for accelerometer were ± 1.5 g to ± 16 g,
271 gyroscope ± 500 °/s to ± 2000 °/s, magnetometer ± 1200 μ T or 1.2 to 4 Ga. Individual sensor sample
272 rates ranged from 10 Hz to 1000 Hz for accelerometers, 10 Hz to 500 Hz for gyroscopes and 50 Hz
273 to 500 Hz for magnetometers.

274

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

276

277 **3.3 Vision capture specification**

278 Several experimental set-ups and specifications were reported in the total **23 vision-based studies**
279 (Table 4). Modality was predominately red, green, blue (RGB) cameras. Depth cameras were utilised
280 (Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015; Kasiri, Fookes, Sridharan, &
281 Morgan, 2017; Reily et al., 2017), which add depth perception for 3-dimensional image mapping.
282 Seven studies clearly reported the use of a single camera set-up (Couceiro, Dias, Mendes, & Araújo,
283 2013; Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014; Hachaj, Ogiela, & Koptyra, 2015;
284 Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Nibali et al., 2017; Reily et al., 2017). One study
285 reported 16 stationary positioned cameras at a ‘bird’s eye view’ (Montoliu et al., 2015), and Ó
286 Conaire et al. (2010) reported the use of one overhead and 8 stationary cameras around a tennis court
287 baseline, although data from two cameras were only used in final analysis due to occlusion issues.
288 Sample frequency and, or pixel resolution were reported in seven of the studies (Couceiro et al.,
289 2013; Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Montoliu et al., 2015;
290 Victor et al., 2017; Zhu et al., 2006), with sample frequencies ranging from 30 Hz to 210 Hz.

291

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

293

294 **3.4 Inertial measurement unit recognition model development methods**

295 Key stages of model development from data pre-processing to recognition techniques for IMU-based
296 studies are presented in Table 5. Data pre-processing filters were reported as either a low-pass filter
297 (n = 7) (Adelsberger & Tröster, 2013; Buckley et al., 2017; Kelly, Coughlan, Green, & Caulfield,

298 2012; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2015, 2017; Rindal, Seeberg, Tjønnås,
299 Haugnes, & Sandbakk, 2018), high-pass filter ($n = 2$) (Kautz et al., 2017; Schuldhaus et al., 2015),
300 or calibration with a filter (Salman et al., 2017). Processing methods were reported in 67% of the
301 IMU-based studies (Adelsberger & Tröster, 2013; Anand, Sharma, Srivastava, Kaligounder, &
302 Prakash, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe, Blanke, Capkevics, & Tröster, 2016;
303 Groh et al., 2016; Groh, Fleckenstein, Kautz, & Eskofier, 2017; Groh, Kautz, & Schuldhaus, 2015;
304 Jensen et al., 2016, 2015; Jiao, Wu, Bie, Umek, & Kos, 2018; Kautz et al., 2017; Kobsar et al., 2014;
305 M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Ó Conaire et al., 2010; Pernek, Kurillo, Stiglic,
306 & Bajcsy, 2015; Qaisar et al., 2013; Salman et al., 2017; Schuldhaus et al., 2015). Methods included,
307 calibration of data (Groh et al., 2016, 2017; Jensen et al., 2015; Qaisar et al., 2013), a one-second
308 window centred around identified activity peaks in the signal (Adelsberger & Tröster, 2013;
309 Schuldhaus et al., 2015), temporal alignment (Pernek et al., 2015), normalisation (Ó Conaire et al.,
310 2010), outlier adjustment (Kobsar et al., 2014) or removal (Salman et al., 2017), and sliding windows
311 ranging from one to 3.5 seconds across the data (Jensen et al., 2016). The three studies that
312 investigated trick classification in skateboarding (Groh et al., 2017, 2015) and snowboarding (Groh
313 et al., 2016) corrected data for different rider board stance styles, termed Regular or Goofy, by
314 inverting signal axes.

315 Movement detection methods were specifically reported in 16 studies (Adelsberger &
316 Tröster, 2013; Anand et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015, Jensen et
317 al., 2013, 2015; Kautz et al., 2017; Kelly et al., 2012; Kos & Kramberger, 2017; Ó Conaire et al.,
318 2010; Rindal et al., 2018; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside, Cant, Connolly,
319 & Reid, 2017). Detection methods included thresholding ($n = 5$), windowing segmenting ($n = 4$), and
320 a combination of threshold and windowing techniques ($n = 5$).

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

326 Algorithms trialled for movement recognition were diverse across the literature (Table 5).
327 Supervised classification using a kernel form of Support Vector Machine (SVM) was most prevalent

328 (n = 16) (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Brock et al., 2017; Buckley et al., 2017;
329 Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Kautz et al., 2017; Kelly et al.,
330 2012; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015;
331 Whiteside et al., 2017). The next highest tested were Naïve Bayesian (NB) (n = 8) (Buckley et al.,
332 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Salman et al., 2017;
333 Schuldhaus et al., 2015) and k-Nearest Neighbour (kNN) (n = 8) (Buckley et al., 2017; Groh et al.,
334 2016, 2017, 2015; Kautz et al., 2017; Ó Conaire et al., 2010; Salman et al., 2017; Whiteside et al.,
335 2017), followed by Random Forests (RF) (n = 7) (Buckley et al., 2017; Groh et al., 2017; Kautz et
336 al., 2017; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Salman et al., 2017; Whiteside et
337 al., 2017). Supervised learning algorithms were the most common (n = 29). One study used an
338 unsupervised discriminative analysis approach for detection and classification of tennis strokes (Kos
339 & Kramberger, 2017). Five IMU-based study investigated a deep learning approach including using
340 Convolutional Neural Networks (CNN) (Anand et al., 2017; Brock et al., 2017; Jiao et al., 2018;
341 Kautz et al., 2017; Rassem et al., 2017) and Long Short Term Memory (LSTM) (Hochreiter &
342 Schmidhuber, 1997) architectures (Rassem et al., 2017; Sharma, Srivastava, Anand, Prakash, &
343 Kaligounder, 2017). In order to assess the effectiveness of the various classifiers from each study,
344 model performance measures quantify and visualise the predictive performance as reported in the
345 following section.

346

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

348

349 **3.5 Inertial measurement unit recognition model evaluation**

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

358 Schuldhaus et al., 2015; Srivastava et al., 2015; Whiteside et al., 2017) and $\geq 80\%$ to 90% ($n = 7$)
359 (Brock & Ohgi, 2017; Brock et al., 2017; Groh et al., 2017; Jensen et al., 2016; M. O'Reilly et al.,
360 2015, 2017; Salman et al., 2017). As an estimate of the generalised performance of a trained model
361 on $n - x$ samples, a form of leave-one-out cross validation (LOO-CV) was used in 47% of studies
362 (Buthe et al., 2016; Groh et al., 2016, 2017, 2015, Jensen et al., 2016, 2013; Kobsar et al., 2014; M.
363 O'Reilly et al., 2015, 2017; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017;
364 Schuldhaus et al., 2015). Precision, specificity and sensitivity (also referred to as recall) evaluations
365 were derived for detection ($n = 6$) and classification models ($n = 10$). Visualisation of prediction
366 results in the form of a confusion matrix featured in six studies (Buthe et al., 2016; Groh et al., 2017;
367 Kautz et al., 2017; Pernek et al., 2015; Rindal et al., 2018; Whiteside et al., 2017).

368

369 *** Table 6 near here***

370

371 3.6 Vision recognition model development methods

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

376 Both machine ($n = 16$) and deep learning ($n = 7$) algorithms were used to recognise
377 movements from vision data. Of these, a kernel form of the SVM algorithm was most common in
378 the studies ($n = 10$) (Couceiro et al., 2013; Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri
379 et al., 2017; Li et al., 2018; Montoliu et al., 2015; M. A. O'Reilly, Whelan, Ward, Delahunt, &
380 Caulfield, 2017b; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Other
381 algorithms included kNN ($n = 3$) (Díaz-Pereira et al., 2014; Montoliu et al., 2015; Ó Conaire et al.,
382 2010), decision tree (DT) ($n = 2$) (Kapela et al., 2015; Liao et al., 2003), RF ($n = 2$) (Kasiri-Bidhendi
383 et al., 2015; Kasiri et al., 2017), and Multilayer Perceptron (MLP) ($n = 2$) (Kapela et al., 2015;
384 Montoliu et al., 2015). Deep learning was investigated in seven studies (Bertasius et al., 2017;
385 Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016; Karpathy et al., 2014a; Nibali et al., 2017;

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

388

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

390

391 **3.7 Vision recognition model evaluation**

392 Performance evaluation methods and results for vision-based studies are reported in Table 8. As with
393 IMU-based studies, classification accuracy was the common method for model evaluations, **featured**
394 **in 61%**. Classification accuracies were reported between 60.9% (Karpathy et al., 2014a) and 100%
395 (Hachaj et al., 2015; Nibali et al., 2017). In grouping the reported highest accuracies for a specific
396 movement that were $\geq 90\%$ ($n = 9$) (Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al.,
397 2017; Li et al., 2018; Montoliu et al., 2015; Nibali et al., 2017; Ó Conaire et al., 2010; Reily et al.,
398 2017; Shah et al., 2007), and $\geq 80\%$ to 90% ($n = 2$) (Horton et al., 2014; Yao & Fei-Fei, 2010). A
399 confusion matrix as a visualisation of model prediction results was used in **nine studies** (Couceiro et
400 al., 2013; Hachaj et al., 2015; Ibrahim et al., 2016; Karpathy et al., 2014a; Kasiri-Bidhendi et al.,
401 2015; Kasiri et al., 2017; Lu, Okuma, & Little, 2009; Shah et al., 2007; Tora et al., 2017). Two
402 studies assessed and reported their model computational average speed (Lu et al., 2009) and time
403 (Reily et al., 2017).

404

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

406

407 **4 Discussion**

408

409 The aim of this systematic review was to evaluate the use of machine and deep learning for sport-
410 specific movement recognition from IMUs and, or computer vision data inputs. Overall, the search
411 yielded **52 studies**, categorised as 29 which used IMUs, **22 vision-based** and one study using both
412 IMUs and vision. Automation or semi-automated sport movement recognition models working in
413 near-real time is of particular interest to avoid the error, cost and time associated with manual
414 methods. Evident in the literature, models are trending towards the potential to provide optimised

415 objective assessments of athletic movement for technical and tactical evaluations. The majority of
416 studies achieved favourable movement recognition results for the main characterising actions of a
417 sport, with several studies exploring further applications such as an automated skill quality evaluation
418 or judgement scoring, for example automated ski jump error evaluation (Brock et al., 2017).

419 Experimental set-up of IMU placement and numbers assigned per participant varied between
420 sporting actions. The sensor attachment locations set by researchers appeared dependent upon the
421 specific sporting conditions and movements, presumably to gain optimal signal data. Proper fixation
422 and alignment of the sensor axes with limb anatomical axes is important in reducing signal error
423 (Fong & Chan, 2010). The attachment site hence requires a biomechanical basis for accuracy of the
424 movement being targeted to obtain reliable data. Single or multiple sensor use per person also
425 impacts model development trade-off between accuracy, analysis complexity, and computational
426 speed or demands. In tennis studies, specificity whilst using a single sensor was demonstrated by
427 mounting the IMU on the wrist or forearm of the racquet arm (Connaghan et al., 2011; Kos &
428 Kramberger, 2017; Srivastava et al., 2015; Whiteside et al., 2017). A single sensor may also be
429 mounted in a low-profile manner on sporting equipment (Groh et al., 2016, 2017, 2015; Jensen et
430 al., 2015). Unobtrusive use of a single IMU to capture generalised movements across the whole body
431 was demonstrated, with an IMU mounted on the posterior head in swimming (Jensen et al., 2016,
432 2013), lower back during running (Kobsar et al., 2014), and between the shoulder blades in rugby
433 union (Kelly et al., 2012).

434 The majority of vision-based studies opted for a single camera set-up of RGB modality. Data
435 output from a single camera as opposed to multiple minimises the volume of data to process,
436 therefore reducing computational effort. However, detailed features may go uncaptured, particularly
437 in team sport competition which consists of multiple individuals participating in the capture space at
438 one time. In contrast, a multiple camera set-up reduces limitations including occlusion and viewpoint
439 variations. However, this may also increase the complexity of the processing and model
440 computational stages. Therefore, a trade-off between computational demands and movement
441 recording accuracy often needs to be made. As stated earlier, the placement of cameras needs to suit
442 the biomechanical nature of the targeted movement and the environment situated in. Common
443 camera capture systems used in sports science research such as Vicon Nexus (Oxford, UK) and
444 OptiTrack (Oregon, USA) were not present in this review. As this review targeted studies

445 investigating during on-field or in-situation sporting contexts, efficiency in data collection is key for
446 routine applications in training and competition. A simple portable RGB camera is easy to set-up in
447 a dynamic and changing environment, such as different soccer pitches, rather than a multiple capture
448 system such as Vicon that requires calibrated precision and are substantially more expensive.

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

461 Supervised learning approaches were dominant across IMU and vision-based studies. This
462 is a method which involves a labelled ground truth training dataset typically manually annotated by
463 sport analysts. Labelled data instances were recorded as up to 15, 000 for vision-based (Victor et al.,
464 2017) and 416, 737 for sensor-based (Rassem et al., 2017) studies. Generation of a training data set
465 for supervised learning can be a tedious and labour-intensive task. It is further complicated if multiple
466 sensors or cameras are incorporated for several targeted movements. A semi-supervised or
467 unsupervised learning approach may be advantageous as data labelling is minimal or not required,
468 potentially reducing human errors in annotation. An unsupervised approach could suit specific
469 problems to explain key data features, via clustering (Mohammed et al., 2016; Sze et al., 2017).
470 Results computed by an unsupervised model (Kos, Ženko, Vlaj, & Kramberger, 2016) for tennis
471 serve, forehand and backhand stroke classification compared favourably well against a proposed
472 supervised approach (Connaghan et al., 2011).

473 Recognition of sport-specific movements was primarily achieved using conventional
474 machine learning approaches, however nine studies implemented deep learning algorithms. It is

475 expected that future model developments will progressively feature deep learning approaches due to
476 development of better hardware, and the advantages of more efficient model learning on large data
477 inputs (Sze et al., 2017). Convolutional Neural networks (CNN) (LeCun, Bottou, Bengio, & Haffner,
478 1998) were the core structure of five of the seven deep learning study models. Briefly, convolution
479 applies several filters, known as kernels, to automatically extract features from raw data inputs. This
480 process works under four key ideas to achieve optimised results: local connection, shared weights,
481 pooling and applying several layers (LeCun et al., 2015; J. B. Yang et al., 2015). Machine learning
482 classifiers modelled with generic hand-crafted features, were compared against a CNN for
483 classifying nine beach volleyball actions using IMUs (Kautz et al., 2017). Unsatisfactory results were
484 obtained from the machine learning model, and the CNN markedly achieved higher classification
485 accuracies (Kautz et al., 2017). The CNN model produced the shortest overall computation times,
486 requiring less computational effort on the same hardware (Kautz et al., 2017). Vision-based CNN
487 models have also shown favourable results when compared to a machine learning study baseline
488 (Karpathy et al., 2014a; Nibali et al., 2017; Victor et al., 2017). Specifically, consistency between a
489 swim stroke detection model for continuous videos in swimming which was then applied to tennis
490 strokes with no domain-specific settings introduced (Victor et al., 2017). The authors of this training
491 approach (Victor et al., 2017) anticipate that this could be applied to train separate models for other
492 sports movement detection as the CNN model demonstrated the ability to learn to process continuous
493 videos into a 1-D signal with the signal peaks corresponding to arbitrary events. General human
494 activity recognition using CNN have shown to be a superior approach over conventional machine
495 learning algorithms using both IMUs (Ravi et al., 2016; J. B. Yang et al., 2015; Zebin et al., 2016;
496 Zeng et al., 2014; Zheng, Liu, Chen, Ge, & Zhao, 2014) and computer vision (Ji et al., 2013;
497 Krizhevsky et al., 2012; LeCun et al., 2015). As machine learning algorithms extract heuristic
498 features requiring domain knowledge, this creates shallower features which can make it harder to
499 infer high-level and context aware activities (J. B. Yang et al., 2015). Given the previously described
500 advantages of deep learning algorithms which apply to CNN, and the recent results of deep learning,
501 future model developments may benefit from exploring these methods in comparison to current
502 bench mark models.

503 Model performance outcome metrics quantify and visualise the error rate between the
504 predicted outcome and true measure. Comparatively, a kernel form of an SVM was the most common

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

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

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

561

562 **5 Conclusions**

563

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

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

590

591 **Acknowledgements**

592 Not applicable.

593

594 **Disclosure statement**

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

597

598 **Funding**

599 The authors received no specific funding for this work.

600

601 **References**

602

- 603 Adelsberger, R., & Tröster, G. (2013). Experts lift differently: Classification of weight-lifting
604 athletes. In *2013 IEEE International Conference on Body Sensor Networks* (pp. 1–6).
605 Cambridge, MA: Body Sensor Networks (BSN). <https://doi.org/10.1109/BSN.2013.6575458>
- 606 Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern*
607 *Recognition Letters*, *48*, 70–80. <https://doi.org/10.1016/j.patrec.2014.04.011>
- 608 Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2017). Wearable motion
609 sensor based analysis of swing sports. In *2017 16th IEEE International Conference on*
610 *Machine Learning and Applications (ICMLA)* (pp. 261–267).
611 <https://doi.org/10.1109/ICMLA.2017.0-149>
- 612 Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports*
613 *Medicine*, *38*(12), 1025–1043. <https://doi.org/10.2165/00007256-200838120-00006>
- 614 Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer*
615 *Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
616 *Bioinformatics)*, *7978 LNAI*, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1
- 617 Bertasius, G., Park, H. S., Yu, S. X., & Shi, J. (2017). Am I a baller? Basketball performance
618 assessment from first-person videos. *Proceedings of the IEEE International Conference on*
619 *Computer Vision*, 2196–2204. <https://doi.org/10.1109/ICCV.2017.239>
- 620 Brock, H., & Ohgi, Y. (2017). Assessing motion style errors in ski jumping using inertial sensor
621 devices. *IEEE Sensors Journal*, (99), 1–11. <https://doi.org/10.1109/JSEN.2017.2699162>
- 622 Brock, H., Ohgi, Y., & Lee, J. (2017). Learning to judge like a human: convolutional networks for
623 classification of ski jumping errors. *Proceedings of the 2017 ACM International Symposium*
624 *on Wearable Computers - ISWC '17*, 106–113. <https://doi.org/10.1145/3123021.3123038>
- 625 Buckley, C., O'Reilly, M. A., Whelan, D., Vallely Farrell, A., Clark, L., Longo, V., ... Caulfield,
626 B. (2017). Binary classification of running fatigue using a single inertial measurement unit. In
627 *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor*
628 *Networks* (pp. 197–201). IEEE. <https://doi.org/10.1109/BSN.2017.7936040>
- 629 Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-
630 worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33.
631 <https://doi.org/http://dx.doi.org/10.1145/2499621>
- 632 Buthe, L., Blanke, U., Capkevics, H., & Tröster, G. (2016). A wearable sensing system for timing
633 analysis in tennis. In *BSN 2016 - 13th Annual Body Sensor Networks Conference* (pp. 43–48).
634 San Francisco, CA. <https://doi.org/10.1109/BSN.2016.7516230>
- 635 Bux, A., Angelov, P., & Habib, Z. (2017). Vision based human activity recognition: A review. In
636 P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in Computational Intelligence*
637 *Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence*
638 (pp. 341–371). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46562-3_23
- 639
- 640 Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozi, G. (2018). Trends supporting the in-field
641 use of wearable inertial sensors for sport performance evaluation: a systematic review.
642 *Sensors*, *18*(3), 873. <https://doi.org/10.3390/s18030873>
- 643 Chambers, R., Gabbett, T., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to
644 quantify sport-specific movements. *Sports Medicine*, *45*(7), 1065–1081.
645 <https://doi.org/10.1007/s40279-015-0332-9>
- 646 Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., Walsh, M., & O'Mathuna, C. (2011).
647 Multi-sensor classification of tennis strokes. *Journal of IEEE Sensors*, 1437–1440.
- 648 Couceiro, M. S., Dias, G., Mendes, R., & Araújo, D. (2013). Accuracy of pattern detection

- 649 methods in the performance of golf putting. *Journal of Motor Behavior*, 45(1), 37–53.
650 <https://doi.org/10.1080/00222895.2012.740100>
- 651 Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic
652 recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement*
653 *Science*, 34(1), 63–80. <https://doi.org/10.1016/j.humov.2014.01.001>
- 654 Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for
655 context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7),
656 645–662. <https://doi.org/10.1007/s00779-010-0293-9>
- 657 Fong, D. T.-P., & Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower
658 limb biomechanics studies: A systematic review. *Sensors*, 10(12), 11556–11565.
659 <https://doi.org/10.3390/s101211556>
- 660 Gabbett, T., Jenkins, D., & Abernethy, B. (2012). Physical demands of professional rugby league
661 training and competition using microtechnology. *Journal of Science and Medicine in Sport*,
662 15, 80–86. <https://doi.org/10.1016/j.jsams.2011.07.004>
- 663 Gabbett, T., Jenkins, D. G., & Abernethy, B. (2011). Physical collisions and injury in professional
664 rugby league match-play. *Journal of Science and Medicine in Sport*, 14, 210–215.
665 <https://doi.org/10.1016/j.jsams.2011.01.002>
- 666 Gatin, P. B., McLean, O. C., Breed, R. V., & Spittle, M. (2014). Tackle and impact detection in
667 elite Australian football using wearable microsensor technology. *Journal of Sports Sciences*,
668 32(10), 947–953. <https://doi.org/10.1080/02640414.2013.868920>
- 669 Gatin, P. B., McLean, O. C., Spittle, M., & Breed, R. V. (2013). Quantification of tackling
670 demands in professional Australian football using integrated wearable athlete tracking
671 technology. *Journal of Science and Medicine in Sport*, 16(6), 589–593.
672 <https://doi.org/10.1016/j.jsams.2013.01.007>
- 673 Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes
674 using principal component analysis. *Journal of Sports Sciences*, 36(2), 229–237.
675 <https://doi.org/10.1080/02640414.2017.1298826>
- 676 Groh, B. H., Fleckenstein, M., & Eskofier, B. M. (2016). Wearable trick classification in freestyle
677 snowboarding. In *13th International Conference on Wearable and Implantable Body Sensor*
678 *Networks (BSN)* (pp. 89–93). IEEE. <https://doi.org/10.1109/BSN.2016.7516238>
- 679 Groh, B. H., Fleckenstein, M., Kautz, T., & Eskofier, B. M. (2017). Classification and visualization
680 of skateboard tricks using wearable sensors. *Pervasive and Mobile Computing*, 40, 42–55.
681 <https://doi.org/10.1016/j.pmcj.2017.05.007>
- 682 Groh, B. H., Kautz, T., & Schuldhuis, D. (2015). IMU-based trick classification in skateboarding.
683 In *KDD Workshop on Large-Scale Sports Analytics*.
- 684 Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods
685 to Oyama karate techniques recognition. *Symmetry*, 7(4), 1670–1698.
686 <https://doi.org/10.3390/sym7041670>
- 687 Hafer, J. F., & Boyer, K. A. (2017). Variability of segment coordination using a vector coding
688 technique: reliability analysis for treadmill walking and running. *Gait and Posture*, 51, 222–
689 227. <https://doi.org/10.1016/j.gaitpost.2016.11.004>
- 690 Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Proceedings Of The*
691 *International Joint Conference On Neural Networks*, 1, 593–605.
692 <https://doi.org/10.1109/IJCNN.1989.118638>
- 693 Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, 9(8),
694 1–32. <https://doi.org/10.1162/neco.1997.9.8.1735>
- 695 Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2014). Classification of passes in
696 football matches using spatiotemporal data. *ArXiv Preprint ArXiv:1407.5093*.
697 <https://doi.org/10.1145/3105576>
- 698 Howe, S. T., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cavanagh, B. P. (2017).
699 Quantifying important differences in athlete movement during collision-based team sports:
700 Accelerometers outperform global positioning systems. In *2017 IEEE International*
701 *Symposium on Inertial Sensors and Systems* (pp. 1–4). Kauai, HI, USA: IEEE.
702 <https://doi.org/10.1109/ISISS.2017.7935655>
- 703 Hulin, B. T., Gabbett, T., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can
704 accurately identify collision events during professional rugby league match-play. *Journal of*
705 *Science and Medicine in Sport*, 20(7), 638–642.
706 <https://doi.org/http://dx.doi.org/10.1016/j.jsams.2016.11.006>
- 707 Ibrahim, M., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A Hierarchical Deep

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

710 Jensen, U., Blank, P., Kugler, P., & Eskofier, B. M. (2016). Unobtrusive and energy-efficient
711 swimming exercise tracking using on-node processing. *IEEE Sensors Journal*, *16*(10), 3972–
712 3980. <https://doi.org/10.1109/JSEN.2016.2530019>

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

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

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

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

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

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

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

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

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

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

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

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

750 Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions
751 in elite level rugby union using a wearable sensing device. *Sports Engineering*, *15*(2), 81–92.
752 Retrieved from <https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5>
753 0088-5

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

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

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

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

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

767 1105. <https://doi.org/http://dx.doi.org/10.1016/j.protecy.2014.09.007>

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

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

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

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

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

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

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

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

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

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

795
796 McNamara, D. J., Gabbett, T., Chapman, P., Naughton, G., & Farhart, P. (2015). The validity of
797 microsensors to automatically detect bowling events and counts in cricket fast bowlers.
798 *International Journal of Sports Physiology and Performance*, 10(1), 71–75.
799 <https://doi.org/10.1123/ijsp.2014-0062>

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

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

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

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

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

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

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

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

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

- 826 *Strength and Conditioning Research*, 31(6), 1726–1736.
- 827 O'Reilly, M., Caulfield, B., Ward, T., Johnston, W., & Doherty, C. (2018). Wearable Inertial
828 Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review.
829 *Sports Medicine*. <https://doi.org/10.1007/s40279-018-0878-4>
- 830 O'Reilly, M., Whelan, D., Chaniyalidis, C., Friel, N., Delahunt, E., Ward, T., & Caulfield, B.
831 (2015). Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE*
832 *12th International Conference on Wearable and Implantable Body Sensor Networks*. IEEE.
833 <https://doi.org/10.1109/BSN.2015.7299380>
- 834 O'Reilly, M., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017). Classification of
835 deadlift biomechanics with wearable inertial measurement units. *Journal of Biomechanics*,
836 58, 155–161. <https://doi.org/10.1080/14763141.2017.1314544>
- 837 Ó Conaire, C., Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., & Buckley, J. (2010).
838 Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings*
839 *of the first ACM international workshop on Analysis and retrieval of tracked events and*
840 *motion in imagery streams* (pp. 51–56). ACM. <https://doi.org/10.1145/1877868.1877882>
- 841 Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength
842 training exercises with wearable sensors. *Journal of Biomedical Informatics*, 58, 145–155.
843 <https://doi.org/10.1016/j.jbi.2015.09.020>
- 844 Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in
845 ubiquitous computing. *International Joint Conference on Artificial Intelligence (IJCAI)*,
846 1729.
- 847 Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision*
848 *Computing*, 28(6), 976–990. <https://doi.org/10.1016/j.imavis.2009.11.014>
- 849 Preece, S. J., Goulermas, J. Y., Kenney, L., & Howard, D. (2009). A comparison of feature
850 extraction methods for the classification of dynamic activities from accelerometer data. *IEEE*
851 *Transactions on Biomedical Engineering*, 56(3), 871–879.
852 <https://doi.org/10.1109/TBME.2008.2006190>
- 853 Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009).
854 Activity identification using body-mounted sensors: A review of classification techniques.
855 *Physiological Measurement*, 30(4), R1–R33. <https://doi.org/10.1088/0967-3334/30/4/R01>
- 856 Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine*
857 *Learning*, 42(3), 203–231. <https://doi.org/10.1023/A:1007601015854>
- 858 Qaisar, S., Imtiaz, S., Glazier, P., Farooq, F., Jamal, A., Iqbal, W., & Lee, S. (2013). A method for
859 cricket bowling action classification and analysis using a system of inertial sensors. In
860 *International Conference on Computational Science and its Applications* (pp. 396–412).
861 Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-39649-6>
- 862 Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2015).
863 Detecting events and key actors in multi-person videos.
864 <https://doi.org/10.1109/CVPR.2016.332>
- 865 Rassem, A., El-Beltagy, M., & Saleh, M. (2017). Cross-country skiing gears classification using
866 deep learning. *ArXiv Preprint ArXiv:1706.08924*. Retrieved from
867 <https://arxiv.org/pdf/1706.08924v1.pdf>
- 868 Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor
869 data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health*
870 *Informatics*, 21(1), 1–1. <https://doi.org/10.1109/JBHI.2016.2633287>
- 871 Reily, B., Zhang, H., & Hoff, W. (2017). Real-time gymnast detection and performance analysis
872 with a portable 3D camera. *Computer Vision and Image Understanding*, 159, 154–163.
873 <https://doi.org/10.1016/j.cviu.2016.11.006>
- 874 Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic
875 classification of sub-techniques in classical cross-country skiing using a machine learning
876 algorithm on micro-sensor data. *Sensors (Switzerland)*, 18(1), 75.
877 <https://doi.org/10.3390/s18010075>
- 878 Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using
879 deep learning neural networks. *Expert Systems with Applications*, 59, 235–244.
880 <https://doi.org/10.1016/j.eswa.2016.04.032>
- 881 Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events
882 in soccer videos. *Applied Artificial Intelligence*, 27(7), 656–667.
883 <https://doi.org/10.1080/08839514.2013.787779>
- 884 Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling

- 885 action in the game of cricket. *Data Mining and Knowledge Discovery*, 31(6), 1706–1734.
886 <https://doi.org/10.1007/s10618-017-0511-4>
- 887 Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B. M. (2015). Inertial
888 sensor-based approach for shot/ pass classification during a soccer match. In *Proc. 21st ACM*
889 *KDD Workshop on Large-Scale Sports Analytics* (pp. 1–4). Sydney, Australia.
- 890 Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving
891 classification performance when training data is skewed. In *9th International Conference on*
892 *Pattern Recognition* (pp. 1–4). <https://doi.org/10.1109/ICPR.2008.4761297>
- 893 Shah, H., Chokalingam, P., Paluri, B., & Pradeep, N. (2007). Automated stroke classification in
894 tennis. *Image Analysis and Recognition*, 1128–1137.
- 895 Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to*
896 *algorithms*. New York, USA: Cambridge University Press.
- 897 Sharma, M., Srivastava, R., Anand, A., Prakash, D., & Kaligounder, L. (2017). Wearable motion
898 sensor based phasic analysis of tennis serve for performance feedback. In *2017 IEEE*
899 *International Conference on Acoustics, Speech and Signal Processing* (pp. 5945–5949). New
900 Orleans, LA: IEEE.
- 901 Sprager, S., & Juric, M. B. (2015). *Inertial sensor-based gait recognition: A review. Sensors*
902 *(Switzerland)* (Vol. 15). <https://doi.org/10.3390/s150922089>
- 903 Srivastava, R., Patwari, A., Kumar, S., Mishra, G., Kaligounder, L., & Sinha, P. (2015). Efficient
904 characterization of tennis shots and game analysis using wearable sensors data. In *2015 IEEE*
905 *Sensors- Proceedings* (pp. 1–4). Busan. <https://doi.org/10.1109/ICSENS.2015.7370311>
- 906 Stein, M., Janetzko, H., Lamprecht, A., Breikreutz, T., Zimmermann, P., Goldlücke, B., ... Keim,
907 D. A. (2018). Bring it to the pitch: combining video and movement data to enhance team
908 sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 13–22.
909 <https://doi.org/10.1109/TVCG.2017.2745181>
- 910 Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks:
911 A tutorial and survey. *IEEE*, 105(2), 2295–2329. Retrieved from
912 <http://arxiv.org/abs/1703.09039>
- 913 Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports:
914 Current applications and research topics. *Computer Vision and Image Understanding*, 159, 3–
915 18. <https://doi.org/10.1016/j.cviu.2017.04.011>
- 916 Titterton, D. H., & Weston, J. L. (2009). *Strapdown inertial navigation technology* (2nd ed.).
917 Reston, VA: AIAA.
- 918 Tora, M. R., Chen, J., & Little, J. J. (2017). Classification of puck possession events in ice hockey.
919 In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*
920 *Workshops* (pp. 147–154). <https://doi.org/10.1109/CVPRW.2017.24>
- 921 Victor, B., He, Z., Morgan, S., & Miniutti, D. (2017). Continuous video to simple signals for
922 swimming stroke detection with convolutional neural networks. *ArXiv Preprint*
923 *ArXiv:1705.09894*. <https://doi.org/10.1111/j.1467-8330.1974.tb00606.x>
- 924 Wagner, D., Kalischewski, K., Velten, J., & Kummert, A. (2017). Activity recognition using
925 inertial sensors and a 2-D convolutional neural network. In IEEE (Ed.), *2017 10th*
926 *International Workshop on Multidimensional (nD) Systems (nDS)* (pp. 1–6).
927 <https://doi.org/10.1109/NDS.2017.8070615>
- 928 Wagner, J. F. (2018). About motion measurement in sports based on gyroscopes and
929 accelerometers - an engineering point of view. *Gyroscopy and Navigation*, 9(1), 1–18.
930 <https://doi.org/10.1134/S2075108718010091>
- 931 Ward, J. A., Lukowicz, P., & Gellersen, H.-W. (2011). Performance metrics for activity
932 recognition. In *ACM Trans. on Intelligent Systems and Technology* (Vol. 2, pp. 111–132).
- 933 Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring hitting load in tennis using
934 inertial sensors and machine learning. *International Journal of Sports Physiology and*
935 *Performance*, 1–20. <https://doi.org/https://doi.org/10.1123/ijsp.2016-0683>
- 936 Wixted, A., Billing, D. C., & James, D. A. (2010). Validation of trunk mounted inertial sensors for
937 analysing running biomechanics under field conditions, using synchronously collected foot
938 contact data. *Sports Engineering*, 12(4), 207–212. <https://doi.org/10.1007/s12283-010-0043-2>
- 939 Wixted, A., Portus, M., Spratford, W., & James, D. A. (2011). Detection of throwing in cricket
940 using wearable sensors. *Sports Technology*, 4(3–4), 134–140.
941 <https://doi.org/10.1080/19346182.2012.725409>
- 942 Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural*
943 *Computation*, 8(7), 1341–1390. <https://doi.org/10.1162/neco.1996.8.7.1391>

- 944 Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimisation. *IEEE*
945 *Transactions on Evolutionary Computation*, *1*(1), 67–82.
946 <https://doi.org/10.1023/A:1021251113462>
- 947 Wundersitz, D. W., Gastin, P. B., Richter, C., Robertson, S., & Netto, K. J. (2015). Validity of a
948 trunk-mounted accelerometer to assess peak accelerations during walking, jogging and
949 running. *European Journal of Sport Science*, *15*(5), 382–390.
950 <https://doi.org/10.1080/17461391.2014.955131>
- 951 Wundersitz, D. W., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of
952 a trunk-mounted accelerometer to measure peak impacts during team sport movements.
953 *International Journal of Sports Medicine*, *36*(9), 742–746. [https://doi.org/10.1055/s-0035-](https://doi.org/10.1055/s-0035-1547265)
954 [1547265](https://doi.org/10.1055/s-0035-1547265)
- 955 Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015).
956 Classification of team sport activities using a single wearable tracking device. *Journal of*
957 *Biomechanics*, *48*(15), 3975–3981. <https://doi.org/10.1016/j.jbiomech.2015.09.015>
- 958 Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for
959 physical activity monitoring. *Sensors*, *10*(8), 7772–7788. <https://doi.org/10.3390/s100807772>
- 960 Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Shonali, K. (2015). Deep convolutional neural
961 networks on multichannel time series for human activity recognition. In *Proceedings of the*
962 *24th International Conference on Artificial Intelligence* (pp. 3995–4001).
- 963 Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object
964 interaction activities. In *Computer Vision and Pattern Recognition* (pp. 17–24). IEEE.
- 965 Young, C., & Reinkensmeyer, D. J. (2014). Judging complex movement performances for
966 excellence: a principal components analysis-based technique applied to competitive diving.
967 *Human Movement Science*, *36*, 107–122. <https://doi.org/10.1016/j.humov.2014.05.009>
- 968 Yu, G., Jang, Y. J., Kim, J., Kim, J. H., Kim, H. Y., Kim, K., & Panday, S. B. (2016). Potential of
969 IMU sensors in performance analysis of professional alpine skiers. *Sensors (Switzerland)*,
970 *16*(4), 1–21. <https://doi.org/10.3390/s16040463>
- 971 Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human Activity Recognition with Inertial
972 Sensors Using a Deep Learning Approach. *Proc. of IEEE Sensors 2016*, (1), 1–3.
973 <https://doi.org/10.1109/ICSENS.2016.7808590>
- 974 Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014).
975 Convolutional neural networks for human activity recognition using mobile sensors. In
976 *Proceedings of the 6th International Conference on Mobile Computing, Applications and*
977 *Services* (pp. 197–205). <https://doi.org/10.4108/icst.mobicase.2014.257786>
- 978 Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity
979 recognition using vision-based method. *Journal of Healthcare Engineering*, *2017*, 1–31.
980 <https://doi.org/10.1155/2017/3090343>
- 981 Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-
982 channels deep convolutional neural networks. In *International Conference on Web-Age*
983 *Information Management* (pp. 298–310). Springer. [https://doi.org/10.1007/978-3-319-08010-](https://doi.org/10.1007/978-3-319-08010-9_33)
984 [9_33](https://doi.org/10.1007/978-3-319-08010-9_33)
- 985 Zhu, G., Xu, C., Gao, W., & Huang, Q. (2006). Action recognition in broadcast tennis video.
986 *Computer Vision in Human-Computer Interaction*, 89–98.
987 https://doi.org/10.1007/11754336_9
- 988 Ziaeefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review.
989 *Pattern Recognition*, *48*(8), 2329–2345. <https://doi.org/10.1016/j.patcog.2015.03.006>
- 990

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

3
4 *Emily E. Cust^{1, 2*}, Alice J. Sweeting^{1, 2}, Kevin Ball¹ and Sam Robertson^{1, 2}*

5
6 **Author details:**

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

8 ² Western Bulldogs Football Club, Footscray, Melbourne, Australia

9
10 **Author ORCID**

11 Sam Robertson 0000-0002-8330-0011

12 Alice Sweeting 0000-0002-9185-6773

13 Emily Cust 0000-0001-6927-6329

14
15 * Corresponding author:

16 Emily Cust

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

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

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

20
21 **Running title:**

22 Machine and deep learning for sport movement recognition review

23

24 .

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39 **Abstract**

40

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41 Objective assessment of an athlete’s performance is of importance in elite sports to facilitate detailed
42 analysis. The implementation of automated detection and recognition of sport-specific movements
43 overcomes the limitations associated with manual performance analysis methods. The object of this
44 study was to systematically review the literature on machine and deep learning for sport-specific
45 movement recognition using inertial measurement unit (IMU) and, or computer vision data inputs.
46 A search of multiple databases was undertaken. Included studies must have investigated a sport-
47 specific movement and analysed via machine or deep learning methods for model development. A
48 total of 52 studies met the inclusion and exclusion criteria. Data pre-processing, processing, model
49 development and evaluation methods varied across the studies. Model development for movement
50 recognition were predominantly undertaken using supervised classification approaches. A kernel
51 form of the Support Vector Machine algorithm was used in 53% of IMU and 50% of vision-based
52 studies. Twelve studies used a deep learning method as a form of Convolutional Neural Network
53 algorithm and one study also adopted a Long Short Term Memory architecture in their model. The
54 adaptation of experimental set-up, data pre-processing, and model development methods are best
55 considered in relation to the characteristics of the targeted sports movement(s).

56

57

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

58 **Key Words:**

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

61 **1. Introduction**

62

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

78 Human activities detected by inertial sensing devices and computer vision are represented
79 as wave signal features corresponding to specific actions, which can be logged and extracted. Human
80 movement activities are considered hierarchically structured and can be broken down to basic
81 movements. Therefore, the context of signal use, intra-class variability, and inter-class similarity
82 between activities require consideration during experimental set-up and model development.
83 Wearable IMUs contain a combination of accelerometer, gyroscope, and magnetometer sensors
84 measuring along one to three axes. These sensors quantify acceleration, angular velocity, and the
85 direction and orientation of travel respectively (Gastin, McLean, Breed, & Spittle, 2014). These
86 sensors can capture repeated movement patterns during sport training and competitions (Camomilla,
87 Bergamini, Fantozzi, & Vannozzi, 2018; Chambers, Gabbett, Cole, & Beard, 2015; J. F. Wagner,
88 2018). Advantages include being wireless, lightweight and self-contained in operation. Inertial
89 measurement units have been utilised in quantifying physical output and tackling impacts in
90 Australian Rules football (Gastin et al., 2014; Gastin, McLean, Spittle, & Breed, 2013) and rugby

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

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

115 Machine learning algorithms learn from data input for automated model building and
116 perform tasks without being explicitly programmed. The algorithm goal is to output a response
117 function $h\sigma(\bar{x})$ that will predict a ground truth variable y from an input vector of variables \bar{x} . Models
118 are run for classification techniques to predict a target class (Kotsiantis, Zaharakis, & Pintelas, 2007),
119 or regression to predict discrete or continuous values. Models are aimed at finding an optimal set of
120 parameters σ to describe the response function, and then make predictions on unseen unlabelled data

121 input. Within these, model training approaches can generally run as supervised learning,
122 unsupervised learning or semi-supervised learning (Mohammed, Khan, & Bashier, 2016; Sze, Chen,
123 Yang, & Emer, 2017).

124 Processing raw data is limited for conventional machine learning algorithms, as they are
125 unable to effectively be trained on abstract and high-dimensional data that is inconsistent, contains
126 missing values or noisy artefacts (Bux et al., 2017; Kautz, 2017). Consequently, several pre-
127 processing stages are required to create a suitable data form for input into the classifier algorithm
128 (Figo, Diniz, Ferreira, & Cardoso, 2010). Filtering (Figo et al., 2010; Wundersitz, Gastin, Robertson,
129 Davey, & Netto, 2015), window capture durations (Mitchell, Monaghan, & O'Connor, 2013; Preece,
130 Goulermas, Kenney, & Howard, 2009; Wundersitz, Josman, et al., 2015), and signal frequency cut-
131 offs (Wundersitz, Gastin, Richter, Robertson, & Netto, 2015; Wundersitz, Gastin, Robertson, et al.,
132 2015) are common techniques applied prior to data prior to dynamic human movement recognition.
133 Well-established filters for processing motion signal data include the Kalman filter (Kautz, 2017;
134 Titterton & Weston, 2009; D. Wagner, Kalischewski, Velten, & Kummert, 2017) and a Fourier
135 transform filter (Preece, Goulermas, Kenney, Howard, et al., 2009) such as a fast Fourier transform
136 (Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015; Preece, Goulermas, Kenney, &
137 Howard, 2009). Near real-time processing benefits from reducing memory requirements,
138 computational demands, and essential bandwidth during whole model implementation. Signal
139 feature extraction and selection favours faster processing by reducing the signals to the critical
140 features that can discriminate the targeted activities (Bulling et al., 2014). Feature extraction involves
141 identifying the key features that help maximise classifier success, and removing features that have
142 minimal impact in the model (Mannini & Sabatini, 2010). Thus, feature selection involves
143 constructing data representations in subspaces with reduced dimensions. These identified variables
144 are represented in a compact feature variable (Mannini & Sabatini, 2010). Common methods include
145 principal component analysis (PCA) (Gløersen, Myklebust, Hallén, & Federolf, 2018; Young &
146 Reinkensmeyer, 2014), vector coding techniques (Hafer & Boyer, 2017) and empirical cumulative
147 distribution functions (ECDF) (Plötz, Hammerla, & Olivier, 2011). An ECDF approach has been
148 shown to be advantageous over PCA as it derives representations of raw input independent of the
149 absolute data ranges, whereas PCA is known to have reduced performance when the input data is not
150 properly normalised (Plötz et al., 2011). For further detailed information on the acquisition, filtering

151 and analysis of IMU data for sports application and vision-based human activity recognition, see
152 (Kautz, 2017) and (Bux et al., 2017), respectively.

153 Deep learning is a division of machine learning, characterised by deeper neural network
154 model architectures and are inspired by the biological neural networks of the human brain (Bengio,
155 2013; LeCun et al., 2015; Sze et al., 2017). The deeper hierarchical models create a profound
156 architecture of multiple hidden layers based on representative learning with several processing and
157 abstraction layers (Bux et al., 2017; J. B. Yang et al., 2015). These computational models allow data
158 input features to be automatically extracted from raw data and transformed to handle unstructured
159 data, including vision (LeCun et al., 2015; Ravi, Wong, Lo, & Yang, 2016). This direct input avoids
160 several processing steps required in machine learning during training and testing, therefore reducing
161 overall computational times. A current key element within deep learning is backpropagation (Hecht-
162 Nielsen, 1989; LeCun, Bottou, Orr, & Müller, 1998). Backpropagation is a fast and computationally
163 efficient algorithm, using gradient descent, that allows training deep neural networks to be tractable
164 (Sze et al., 2017). Human activity recognition has mainly been performed using conventional
165 machine learning classifiers. Recently, deep learning techniques have enhanced the bench mark and
166 applications for IMUs (Kautz et al., 2017; Ravi et al., 2016; Ronao & Cho, 2016; J. B. Yang et al.,
167 2015; Zebin, Scully, & Ozanyan, 2016; Zeng et al., 2014) and vision (Ji, Yang, Yu, & Xu, 2013;
168 Karpathy et al., 2014a; Krizhevsky, Sutskever, & Hinton, 2012; Nibali, He, Morgan, & Greenwood,
169 2017) in human movement recognition producing more superior model performance accuracy.

170 The objective of this study was to systematically review the literature investigating sport-
171 specific automated movement detection and recognition. The review focusses on the various
172 technologies, analysis techniques and performance outcome measures utilised. There are several
173 reviews within this field that are sensor-based including wearable IMUs for lower limb biomechanics
174 and exercises (Fong & Chan, 2010; M. O'Reilly, Caulfield, Ward, Johnston, & Doherty, 2018),
175 swimming analysis (Magalhaes, Vannozzi, Gatta, & Fantozzi, 2015; Mooney et al., 2015),
176 quantifying sporting movements (Chambers et al., 2015) and physical activity monitoring (C. C.
177 Yang & Hsu, 2010). A recent systematic review has provided an evaluation on the in-field use of
178 inertial-based sensors for various performance evaluation applications (Camomilla et al., 2018).
179 Vision-based methods for human activity recognition (Aggarwal & Xia, 2014; Bux et al., 2017; Ke
180 et al., 2013; Zhang et al., 2017), semantic human activity recognition (Ziaefard & Bergevin, 2015)

181 and motion analysis in sport (Barris & Button, 2008) have also been reviewed. However, to date,
182 there is no systematic review across sport-specific movement detection and recognition via machine
183 or deep learning. Specifically, incorporating IMUs and vision-based data input, focussing on in-field
184 applications as opposed to laboratory-based protocols and detailing the analysis and machine
185 learning methods used.

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

194 **2. Methods**

196 **2.1 Search strategy**

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

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

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

211

212 2.2 Data extraction

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

240 attachment of sensors (IMUs), sample frequency, and sensor data types used in analysis (IMUs).

241 Studies identified and full data extracted were reviewed by a second author.

242

243 3. Results

244

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

250

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

252

253 3.1 Experimental design

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

270 (Srivastava et al., 2015). Semantic descriptors were reported for classification models that predicted
271 athlete training background, experience and fatigue level. These included running (Buckley et al.,
272 2017; Kobsar, Osis, Hettinga, & Ferber, 2014), rating of gymnastic routines (Reily, Zhang, & Hoff,
273 2017), soccer pass classification based on its quality (Horton, Gudmundsson, Chawla, & Estephan,
274 2014), cricket bowling legality (Qaisar et al., 2013; Salman, Qaisar, & Qamar, 2017), ski jump error
275 analysis (Brock & Ohgi, 2017; Brock, Ohgi, & Lee, 2017) and strength training technique deviations
276 (M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a; M. O'Reilly et al., 2015; M.
277 O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017). One approach (Yao & Fei-Fei, 2010),
278 encoded the mutual context of human pose and sporting equipment using semantics, to facilitate the
279 detection and classification of movements including a cricket bat and batsman coupled movements.

280 Total participant numbers for IMU-based studies ranged from one (Qaisar et al., 2013) to 30
281 (Kautz et al., 2017). Reported data individual instance sample sizes for sensor studies ranged from
282 150 (Salman et al., 2017) to 416, 737 (Rassem, El-Beltagy, & Saleh, 2017). Vision-based studies
283 that explicitly reported total participant details ranged from five (Ó Conaire et al., 2010) to 40 (Victor
284 et al., 2017). Vision dataset sample sizes varied across studies, from 50 individual action clips (Liao
285 et al., 2003) to 15, 000 (Victor et al., 2017). One study (Karpathy et al., 2014a) used the publicly
286 available Sports-1M, as previously described. Vision-based studies also reported datasets in total
287 time, 10.3 hours (Bertasius, Park, Yu, & Shi, 2017), 3 hours (Montoliu, Martín-Félez, Torres-
288 Sospedra, & Martínez-Usó, 2015), 1, 500 minutes (Shah et al., 2007), and 50 hours (Kapela et al.,
289 2015), and by frame numbers, 6, 035 frames (Zhu, Xu, Gao, & Huang, 2006) and 10, 115 frames
290 (Reily et al., 2017).

291

292 **3.2 Inertial measurement unit specifications**

293 A range of commercially available and custom-built IMUs were used in the IMU-based studies (n=
294 30), as presented in Table 3. Of these, 23% reported using a custom-built sensor. Of the IMU-based
295 studies, the number of sensors mounted or attached to each participant or sporting equipment piece
296 ranged from one to nine. The majority of studies (n= 22) provided adequate details of sensor
297 specifications including sensor type, axes, measurement range, and sample rate used. At least one
298 characteristic of sensor measurement range or sample rate used in data collection was missing from
299 eight studies. All studies used triaxial sensors and collected accelerometer data. For analysis and

300 model development, individual sensor data consisted of only accelerometer data (n = 8), both
301 accelerometer and gyroscope data (n = 15), and accelerometer, gyroscope and magnetometer data (n
302 = 7). The individual sensor measurement ranges reported for accelerometer were ± 1.5 g to ± 16 g,
303 gyroscope ± 500 °/s to ± 2000 °/s, magnetometer ± 1200 μ T or 1.2 to 4 Ga. Individual sensor sample
304 rates ranged from 10 Hz to 1000 Hz for accelerometers, 10 Hz to 500 Hz for gyroscopes and 50 Hz
305 to 500 Hz for magnetometers.

306

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

308

309 **3.3 Vision capture specification**

310 Several experimental set-ups and specifications were reported in the total **23 vision-based studies**
311 (Table 4). Modality was predominately red, green, blue (RGB) cameras. Depth cameras were utilised
312 (Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015; Kasiri, Fookes, Sridharan, &
313 Morgan, 2017; Reily et al., 2017), which add depth perception for 3-dimensional image mapping.
314 Seven studies clearly reported the use of a single camera set-up (Couceiro, Dias, Mendes, & Araújo,
315 2013; Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014; Hachaj, Ogiela, & Koptyra, 2015;
316 Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Nibali et al., 2017; Reily et al., 2017). One study
317 reported 16 stationary positioned cameras at a ‘bird’s eye view’ (Montoliu et al., 2015), and Ó
318 Conaire et al. (2010) reported the use of one overhead and 8 stationary cameras around a tennis court
319 baseline, although data from two cameras were only used in final analysis due to occlusion issues.
320 Sample frequency and, or pixel resolution were reported in seven of the studies (Couceiro et al.,
321 2013; Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Montoliu et al., 2015;
322 Victor et al., 2017; Zhu et al., 2006), with sample frequencies ranging from 30 Hz to 210 Hz.

323

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

325

326 **3.4 Inertial measurement unit recognition model development methods**

327 Key stages of model development from data pre-processing to recognition techniques for IMU-based
328 studies are presented in Table 5. Data pre-processing filters were reported as either a low-pass filter
329 (n = 7) (Adelsberger & Tröster, 2013; Buckley et al., 2017; Kelly, Coughlan, Green, & Caulfield,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

330 2012; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2015, 2017; Rindal, Seeberg, Tjønnås,
331 Haugnes, & Sandbakk, 2018), high-pass filter ($n = 2$) (Kautz et al., 2017; Schuldhaus et al., 2015),
332 or calibration with a filter (Salman et al., 2017). Processing methods were reported in 67% of the
333 IMU-based studies (Adelsberger & Tröster, 2013; Anand, Sharma, Srivastava, Kaligounder, &
334 Prakash, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe, Blanke, Capkevics, & Tröster, 2016;
335 Groh et al., 2016; Groh, Fleckenstein, Kautz, & Eskofier, 2017; Groh, Kautz, & Schuldhaus, 2015;
336 Jensen et al., 2016, 2015; Jiao, Wu, Bie, Umek, & Kos, 2018; Kautz et al., 2017; Kobsar et al., 2014;
337 M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Ó Conaire et al., 2010; Pernek, Kurillo, Stiglic,
338 & Bajcsy, 2015; Qaisar et al., 2013; Salman et al., 2017; Schuldhaus et al., 2015). Methods included,
339 calibration of data (Groh et al., 2016, 2017; Jensen et al., 2015; Qaisar et al., 2013), a one-second
340 window centred around identified activity peaks in the signal (Adelsberger & Tröster, 2013;
341 Schuldhaus et al., 2015), temporal alignment (Pernek et al., 2015), normalisation (Ó Conaire et al.,
342 2010), outlier adjustment (Kobsar et al., 2014) or removal (Salman et al., 2017), and sliding windows
343 ranging from one to 3.5 seconds across the data (Jensen et al., 2016). The three studies that
344 investigated trick classification in skateboarding (Groh et al., 2017, 2015) and snowboarding (Groh
345 et al., 2016) corrected data for different rider board stance styles, termed Regular or Goofy, by
346 inverting signal axes.

347 Movement detection methods were specifically reported in 16 studies (Adelsberger &
348 Tröster, 2013; Anand et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015, Jensen et
349 al., 2013, 2015; Kautz et al., 2017; Kelly et al., 2012; Kos & Kramberger, 2017; Ó Conaire et al.,
350 2010; Rindal et al., 2018; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside, Cant, Connolly,
351 & Reid, 2017). Detection methods included thresholding ($n = 5$), windowing segmenting ($n = 4$), and
352 a combination of threshold and windowing techniques ($n = 5$).

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

358 Algorithms trialled for movement recognition were diverse across the literature (Table 5).
359 Supervised classification using a kernel form of Support Vector Machine (SVM) was most prevalent

360 (n = 16) (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Brock et al., 2017; Buckley et al., 2017;
361 Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Kautz et al., 2017; Kelly et al.,
362 2012; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015;
363 Whiteside et al., 2017). The next highest tested were Naïve Bayesian (NB) (n = 8) (Buckley et al.,
364 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Salman et al., 2017;
365 Schuldhaus et al., 2015) and k-Nearest Neighbour (kNN) (n = 8) (Buckley et al., 2017; Groh et al.,
366 2016, 2017, 2015; Kautz et al., 2017; Ó Conaire et al., 2010; Salman et al., 2017; Whiteside et al.,
367 2017), followed by Random Forests (RF) (n = 7) (Buckley et al., 2017; Groh et al., 2017; Kautz et
368 al., 2017; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Salman et al., 2017; Whiteside et
369 al., 2017). Supervised learning algorithms were the most common (n = 29). One study used an
370 unsupervised discriminative analysis approach for detection and classification of tennis strokes (Kos
371 & Kramberger, 2017). Five IMU-based study investigated a deep learning approach including using
372 Convolutional Neural Networks (CNN) (Anand et al., 2017; Brock et al., 2017; Jiao et al., 2018;
373 Kautz et al., 2017; Rassem et al., 2017) and Long Short Term Memory (LSTM) (Hochreiter &
374 Schmidhuber, 1997) architectures (Rassem et al., 2017; Sharma, Srivastava, Anand, Prakash, &
375 Kaligounder, 2017). In order to assess the effectiveness of the various classifiers from each study,
376 model performance measures quantify and visualise the predictive performance as reported in the
377 following section.

378

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

380

381 **3.5 Inertial measurement unit recognition model evaluation**

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

390 Schuldhaus et al., 2015; Srivastava et al., 2015; Whiteside et al., 2017) and $\geq 80\%$ to 90% (n = 7)
391 (Brock & Ohgi, 2017; Brock et al., 2017; Groh et al., 2017; Jensen et al., 2016; M. O'Reilly et al.,
392 2015, 2017; Salman et al., 2017). As an estimate of the generalised performance of a trained model
393 on $n - x$ samples, a form of leave-one-out cross validation (LOO-CV) was used in 47% of studies
394 (Buthe et al., 2016; Groh et al., 2016, 2017, 2015, Jensen et al., 2016, 2013; Kobsar et al., 2014; M.
395 O'Reilly et al., 2015, 2017; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017;
396 Schuldhaus et al., 2015). Precision, specificity and sensitivity (also referred to as recall) evaluations
397 were derived for detection (n = 6) and classification models (n = 10). Visualisation of prediction
398 results in the form of a confusion matrix featured in six studies (Buthe et al., 2016; Groh et al., 2017;
399 Kautz et al., 2017; Pernek et al., 2015; Rindal et al., 2018; Whiteside et al., 2017).

400

401 *** Table 6 near here***

402

403 3.6 Vision recognition model development methods

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

408 Both machine (n = 16) and deep learning (n = 7) algorithms were used to recognise
409 movements from vision data. Of these, a kernel form of the SVM algorithm was most common in
410 the studies (n = 10) (Couceiro et al., 2013; Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri
411 et al., 2017; Li et al., 2018; Montoliu et al., 2015; M. A. O'Reilly, Whelan, Ward, Delahunt, &
412 Caulfield, 2017b; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Other
413 algorithms included kNN (n = 3) (Díaz-Pereira et al., 2014; Montoliu et al., 2015; Ó Conaire et al.,
414 2010), decision tree (DT) (n = 2) (Kapela et al., 2015; Liao et al., 2003), RF (n = 2) (Kasiri-Bidhendi
415 et al., 2015; Kasiri et al., 2017), and Multilayer Perceptron (MLP) (n = 2) (Kapela et al., 2015;
416 Montoliu et al., 2015). Deep learning was investigated in seven studies (Bertasius et al., 2017;
417 Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016; Karpathy et al., 2014a; Nibali et al., 2017;

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

420

421 *** Table 7 near here***

422

423 3.7 Vision recognition model evaluation

424 Performance evaluation methods and results for vision-based studies are reported in Table 8. As with
425 IMU-based studies, classification accuracy was the common method for model evaluations, featured
426 in 61%. Classification accuracies were reported between 60.9% (Karpathy et al., 2014a) and 100%
427 (Hachaj et al., 2015; Nibali et al., 2017). In grouping the reported highest accuracies for a specific
428 movement that were $\geq 90\%$ ($n = 9$) (Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al.,
429 2017; Li et al., 2018; Montoliu et al., 2015; Nibali et al., 2017; Ó Conaire et al., 2010; Reily et al.,
430 2017; Shah et al., 2007), and $\geq 80\%$ to 90% ($n = 2$) (Horton et al., 2014; Yao & Fei-Fei, 2010). A
431 confusion matrix as a visualisation of model prediction results was used in nine studies (Couceiro et
432 al., 2013; Hachaj et al., 2015; Ibrahim et al., 2016; Karpathy et al., 2014a; Kasiri-Bidhendi et al.,
433 2015; Kasiri et al., 2017; Lu, Okuma, & Little, 2009; Shah et al., 2007; Tora et al., 2017). Two
434 studies assessed and reported their model computational average speed (Lu et al., 2009) and time
435 (Reily et al., 2017).

436

437 *** Table 8 near here***

438

439 4 Discussion

440

441 The aim of this systematic review was to evaluate the use of machine and deep learning for sport-
442 specific movement recognition from IMUs and, or computer vision data inputs. Overall, the search
443 yielded 52 studies, categorised as 29 which used IMUs, 22 vision-based and one study using both
444 IMUs and vision. Automation or semi-automated sport movement recognition models working in
445 near-real time is of particular interest to avoid the error, cost and time associated with manual
446 methods. Evident in the literature, models are trending towards the potential to provide optimised

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

447 objective assessments of athletic movement for technical and tactical evaluations. The majority of
448 studies achieved favourable movement recognition results for the main characterising actions of a
449 sport, with several studies exploring further applications such as an automated skill quality evaluation
450 or judgement scoring, for example automated ski jump error evaluation (Brock et al., 2017).

451 Experimental set-up of IMU placement and numbers assigned per participant varied between
452 sporting actions. The sensor attachment locations set by researchers appeared dependent upon the
453 specific sporting conditions and movements, presumably to gain optimal signal data. Proper fixation
454 and alignment of the sensor axes with limb anatomical axes is important in reducing signal error
455 (Fong & Chan, 2010). The attachment site hence requires a biomechanical basis for accuracy of the
456 movement being targeted to obtain reliable data. Single or multiple sensor use per person also
457 impacts model development trade-off between accuracy, analysis complexity, and computational
458 speed or demands. In tennis studies, specificity whilst using a single sensor was demonstrated by
459 mounting the IMU on the wrist or forearm of the racquet arm (Connaghan et al., 2011; Kos &
460 Kramberger, 2017; Srivastava et al., 2015; Whiteside et al., 2017). A single sensor may also be
461 mounted in a low-profile manner on sporting equipment (Groh et al., 2016, 2017, 2015; Jensen et
462 al., 2015). Unobtrusive use of a single IMU to capture generalised movements across the whole body
463 was demonstrated, with an IMU mounted on the posterior head in swimming (Jensen et al., 2016,
464 2013), lower back during running (Kobsar et al., 2014), and between the shoulder blades in rugby
465 union (Kelly et al., 2012).

466 The majority of vision-based studies opted for a single camera set-up of RGB modality. Data
467 output from a single camera as opposed to multiple minimises the volume of data to process,
468 therefore reducing computational effort. However, detailed features may go uncaptured, particularly
469 in team sport competition which consists of multiple individuals participating in the capture space at
470 one time. In contrast, a multiple camera set-up reduces limitations including occlusion and viewpoint
471 variations. However, this may also increase the complexity of the processing and model
472 computational stages. Therefore, a trade-off between computational demands and movement
473 recording accuracy often needs to be made. As stated earlier, the placement of cameras needs to suit
474 the biomechanical nature of the targeted movement and the environment situated in. Common
475 camera capture systems used in sports science research such as Vicon Nexus (Oxford, UK) and
476 OptiTrack (Oregon, USA) were not present in this review. As this review targeted studies

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

477 investigating during on-field or in-situation sporting contexts, efficiency in data collection is key for
478 routine applications in training and competition. A simple portable RGB camera is easy to set-up in
479 a dynamic and changing environment, such as different soccer pitches, rather than a multiple capture
480 system such as Vicon that requires calibrated precision and are substantially more expensive.

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

493 Supervised learning approaches were dominant across IMU and vision-based studies. This
494 is a method which involves a labelled ground truth training dataset typically manually annotated by
495 sport analysts. Labelled data instances were recorded as up to 15, 000 for vision-based (Victor et al.,
496 2017) and 416, 737 for sensor-based (Rassem et al., 2017) studies. Generation of a training data set
497 for supervised learning can be a tedious and labour-intensive task. It is further complicated if multiple
498 sensors or cameras are incorporated for several targeted movements. A semi-supervised or
499 unsupervised learning approach may be advantageous as data labelling is minimal or not required,
500 potentially reducing human errors in annotation. An unsupervised approach could suit specific
501 problems to explain key data features, via clustering (Mohammed et al., 2016; Sze et al., 2017).
502 Results computed by an unsupervised model (Kos, Ženko, Vlaj, & Kramberger, 2016) for tennis
503 serve, forehand and backhand stroke classification compared favourably well against a proposed
504 supervised approach (Connaghan et al., 2011).

505 Recognition of sport-specific movements was primarily achieved using conventional
506 machine learning approaches, however nine studies implemented deep learning algorithms. It is

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

507 expected that future model developments will progressively feature deep learning approaches due to
508 development of better hardware, and the advantages of more efficient model learning on large data
509 inputs (Sze et al., 2017). Convolutional Neural networks (CNN) (LeCun, Bottou, Bengio, & Haffner,
510 1998) were the core structure of five of the seven deep learning study models. Briefly, convolution
511 applies several filters, known as kernels, to automatically extract features from raw data inputs. This
512 process works under four key ideas to achieve optimised results: local connection, shared weights,
513 pooling and applying several layers (LeCun et al., 2015; J. B. Yang et al., 2015). Machine learning
514 classifiers modelled with generic hand-crafted features, were compared against a CNN for
515 classifying nine beach volleyball actions using IMUs (Kautz et al., 2017). Unsatisfactory results were
516 obtained from the machine learning model, and the CNN markedly achieved higher classification
517 accuracies (Kautz et al., 2017). The CNN model produced the shortest overall computation times,
518 requiring less computational effort on the same hardware (Kautz et al., 2017). Vision-based CNN
519 models have also shown favourable results when compared to a machine learning study baseline
520 (Karpathy et al., 2014a; Nibali et al., 2017; Victor et al., 2017). Specifically, consistency between a
521 swim stroke detection model for continuous videos in swimming which was then applied to tennis
522 strokes with no domain-specific settings introduced (Victor et al., 2017). The authors of this training
523 approach (Victor et al., 2017) anticipate that this could be applied to train separate models for other
524 sports movement detection as the CNN model demonstrated the ability to learn to process continuous
525 videos into a 1-D signal with the signal peaks corresponding to arbitrary events. General human
526 activity recognition using CNN have shown to be a superior approach over conventional machine
527 learning algorithms using both IMUs (Ravi et al., 2016; J. B. Yang et al., 2015; Zebin et al., 2016;
528 Zeng et al., 2014; Zheng, Liu, Chen, Ge, & Zhao, 2014) and computer vision (Ji et al., 2013;
529 Krizhevsky et al., 2012; LeCun et al., 2015). As machine learning algorithms extract heuristic
530 features requiring domain knowledge, this creates shallower features which can make it harder to
531 infer high-level and context aware activities (J. B. Yang et al., 2015). Given the previously described
532 advantages of deep learning algorithms which apply to CNN, and the recent results of deep learning,
533 future model developments may benefit from exploring these methods in comparison to current
534 bench mark models.

535 Model performance outcome metrics quantify and visualise the error rate between the
536 predicted outcome and true measure. Comparatively, a kernel form of an SVM was the most common

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

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

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

593

594 **5 Conclusions**

595

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
596 This systematic review reported on the literature using machine and deep learning methods to
597 automate sport-specific movement recognition. In addressing the research questions, both IMUs and
598 computer vision have demonstrated capacity in improving the information gained from sport
599 movement and skill recognition for performance analysis. A range of methods for model
600 development were used across the reviewed studies producing varying results. Conventional machine
601 learning algorithms such as Support Vector Machines and Neural Networks were most commonly
602 implemented. Yet in those studies which applied deep learning algorithms such as Convolutional
603 Neural Networks, these methods outperformed the machine learning algorithms in comparison.
604 Typically, the models were evaluated using a leave-one-out cross validation method and reported
605 model performances as a classification accuracy score. Intuitively, the adaptation of experimental
606 set-up, data processing, and recognition methods used are best considered in relation to the
607 characteristics of the sport and targeted movement(s). Consulting current models within or similar to
608 the targeted sport and movement is of benefit to address bench mark model performances and identify
609 areas for improvement. The application within the sporting domain of machine learning and
610 automated sport analysis coding for consistent uniform usage appears currently a challenging
611 prospect, considering the dynamic nature, equipment restrictions and varying environments arising
612 in different sports.

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
613 Future work may look to adopt, adapt and expand on current models associated with a
614 specific sports movement to work towards flexible models for mainstream analysis implementation.
615 Investigation of deep learning methods in comparison to conventional machine learning algorithms
616 would be of particular interest to evaluate if the trend of superior performances is beneficial for sport-
617 specific movement recognition. Analysis as to whether IMUs and vision alone or together yield
618 enhanced results in relation to a specific sport and its implementation efficiency would also be of
619 value. In consideration of the reported study information, this review can assist future researchers in
620 broadening investigative approaches for sports performance analysis as a potential to enhancing upon
621 current methods.

57
58
59
60
61
62
63
64
65
622

623 **Acknowledgements**

624 Not applicable.

625

626 **Disclosure statement**

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

629

630 **Funding**

631 The authors received no specific funding for this work.

632

633 **Author ORCID**

634 Sam Robertson 0000-0002-8330-0011

635 Alice Sweeting 0000-0002-9185-6773

636 Emily Cust 0000-0001-6927-6329

637

638 **References**

639

- 640 Adelsberger, R., & Tröster, G. (2013). Experts lift differently: Classification of weight-lifting
641 athletes. In *2013 IEEE International Conference on Body Sensor Networks* (pp. 1–6).
642 Cambridge, MA: Body Sensor Networks (BSN). <https://doi.org/10.1109/BSN.2013.6575458>
- 643 Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern*
644 *Recognition Letters*, *48*, 70–80. <https://doi.org/10.1016/j.patrec.2014.04.011>
- 645 Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2017). Wearable motion
646 sensor based analysis of swing sports. In *2017 16th IEEE International Conference on*
647 *Machine Learning and Applications (ICMLA)* (pp. 261–267).
648 <https://doi.org/10.1109/ICMLA.2017.0-149>
- 649 Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports*
650 *Medicine*, *38*(12), 1025–1043. <https://doi.org/10.2165/00007256-200838120-00006>
- 651 Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer*
652 *Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
653 *Bioinformatics)*, *7978 LNAI*, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1
- 654 Bertasius, G., Park, H. S., Yu, S. X., & Shi, J. (2017). Am I a baller? Basketball performance
655 assessment from first-person videos. *Proceedings of the IEEE International Conference on*
656 *Computer Vision*, 2196–2204. <https://doi.org/10.1109/ICCV.2017.239>
- 657 Brock, H., & Ohgi, Y. (2017). Assessing motion style errors in ski jumping using inertial sensor
658 devices. *IEEE Sensors Journal*, (99), 1–11. <https://doi.org/10.1109/JSEN.2017.2699162>
- 659 Brock, H., Ohgi, Y., & Lee, J. (2017). Learning to judge like a human: convolutional networks for
660 classification of ski jumping errors. *Proceedings of the 2017 ACM International Symposium*
661 *on Wearable Computers - ISWC '17*, 106–113. <https://doi.org/10.1145/3123021.3123038>
- 662 Buckley, C., O'Reilly, M. A., Whelan, D., Vallely Farrell, A., Clark, L., Longo, V., ... Caulfield,
663 B. (2017). Binary classification of running fatigue using a single inertial measurement unit. In
664 *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor*
665 *Networks* (pp. 197–201). IEEE. <https://doi.org/10.1109/BSN.2017.7936040>
- 666 Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-
667 worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33.
668 <https://doi.org/http://dx.doi.org/10.1145/2499621>
- 669 Buthe, L., Blanke, U., Capkevics, H., & Tröster, G. (2016). A wearable sensing system for timing
670 analysis in tennis. In *BSN 2016 - 13th Annual Body Sensor Networks Conference* (pp. 43–48).
671 San Francisco, CA. <https://doi.org/10.1109/BSN.2016.7516230>
- 672 Bux, A., Angelov, P., & Habib, Z. (2017). Vision based human activity recognition: A review. In
673 P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in Computational Intelligence*
674 *Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence*
675 (pp. 341–371). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46562-3_23
- 676

- 677 Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field
678 use of wearable inertial sensors for sport performance evaluation: a systematic review.
679 *Sensors*, 18(3), 873. <https://doi.org/10.3390/s18030873>
- 680 Chambers, R., Gabbett, T., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to
681 quantify sport-specific movements. *Sports Medicine*, 45(7), 1065–1081.
682 <https://doi.org/10.1007/s40279-015-0332-9>
- 683 Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., Walsh, M., & O'Mathuna, C. (2011).
684 Multi-sensor classification of tennis strokes. *Journal of IEEE Sensors*, 1437–1440.
- 685 Couceiro, M. S., Dias, G., Mendes, R., & Araújo, D. (2013). Accuracy of pattern detection
686 methods in the performance of golf putting. *Journal of Motor Behavior*, 45(1), 37–53.
687 <https://doi.org/10.1080/00222895.2012.740100>
- 688 Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic
689 recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement*
690 *Science*, 34(1), 63–80. <https://doi.org/10.1016/j.humov.2014.01.001>
- 691 Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for
692 context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7),
693 645–662. <https://doi.org/10.1007/s00779-010-0293-9>
- 694 Fong, D. T.-P., & Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower
695 limb biomechanics studies: A systematic review. *Sensors*, 10(12), 11556–11565.
696 <https://doi.org/10.3390/s101211556>
- 697 Gabbett, T., Jenkins, D., & Abernethy, B. (2012). Physical demands of professional rugby league
698 training and competition using microtechnology. *Journal of Science and Medicine in Sport*,
699 15, 80–86. <https://doi.org/10.1016/j.jsams.2011.07.004>
- 700 Gabbett, T., Jenkins, D. G., & Abernethy, B. (2011). Physical collisions and injury in professional
701 rugby league match-play. *Journal of Science and Medicine in Sport*, 14, 210–215.
702 <https://doi.org/10.1016/j.jsams.2011.01.002>
- 703 Gastin, P. B., McLean, O. C., Breed, R. V., & Spittle, M. (2014). Tackle and impact detection in
704 elite Australian football using wearable microsensor technology. *Journal of Sports Sciences*,
705 32(10), 947–953. <https://doi.org/10.1080/02640414.2013.868920>
- 706 Gastin, P. B., McLean, O. C., Spittle, M., & Breed, R. V. (2013). Quantification of tackling
707 demands in professional Australian football using integrated wearable athlete tracking
708 technology. *Journal of Science and Medicine in Sport*, 16(6), 589–593.
709 <https://doi.org/10.1016/j.jsams.2013.01.007>
- 710 Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes
711 using principal component analysis. *Journal of Sports Sciences*, 36(2), 229–237.
712 <https://doi.org/10.1080/02640414.2017.1298826>
- 713 Groh, B. H., Fleckenstein, M., & Eskofier, B. M. (2016). Wearable trick classification in freestyle
714 snowboarding. In *13th International Conference on Wearable and Implantable Body Sensor*
715 *Networks (BSN)* (pp. 89–93). IEEE. <https://doi.org/10.1109/BSN.2016.7516238>
- 716 Groh, B. H., Fleckenstein, M., Kautz, T., & Eskofier, B. M. (2017). Classification and visualization
717 of skateboard tricks using wearable sensors. *Pervasive and Mobile Computing*, 40, 42–55.
718 <https://doi.org/10.1016/j.pmcj.2017.05.007>
- 719 Groh, B. H., Kautz, T., & Schuldhuis, D. (2015). IMU-based trick classification in skateboarding.
720 In *KDD Workshop on Large-Scale Sports Analytics*.
- 721 Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods
722 to Oyama karate techniques recognition. *Symmetry*, 7(4), 1670–1698.
723 <https://doi.org/10.3390/sym7041670>
- 724 Hafer, J. F., & Boyer, K. A. (2017). Variability of segment coordination using a vector coding
725 technique: reliability analysis for treadmill walking and running. *Gait and Posture*, 51, 222–
726 227. <https://doi.org/10.1016/j.gaitpost.2016.11.004>
- 727 Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Proceedings Of The*
728 *International Joint Conference On Neural Networks*, 1, 593–605.
729 <https://doi.org/10.1109/IJCNN.1989.118638>
- 730 Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, 9(8),
731 1–32. <https://doi.org/10.1162/neco.1997.9.8.1735>
- 732 Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2014). Classification of passes in
733 football matches using spatiotemporal data. *ArXiv Preprint ArXiv:1407.5093*.
734 <https://doi.org/10.1145/3105576>
- 735 Howe, S. T., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cavanagh, B. P. (2017).

- 736 Quantifying important differences in athlete movement during collision-based team sports:
737 Accelerometers outperform global positioning systems. In *2017 IEEE International*
738 *Symposium on Inertial Sensors and Systems* (pp. 1–4). Kauai, HI, USA: IEEE.
739 <https://doi.org/10.1109/ISISS.2017.7935655>
- 740 Hulin, B. T., Gabbett, T., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can
741 accurately identify collision events during professional rugby league match-play. *Journal of*
742 *Science and Medicine in Sport*, *20*(7), 638–642.
743 <https://doi.org/http://dx.doi.org/10.1016/j.jsams.2016.11.006>
- 744 Ibrahim, M., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A Hierarchical Deep
745 Temporal Model for Group Activity Recognition. *Cvpr*, 1971–1980.
746 <https://doi.org/10.1109/CVPR.2016.217>
- 747 Jensen, U., Blank, P., Kugler, P., & Eskofier, B. M. (2016). Unobtrusive and energy-efficient
748 swimming exercise tracking using on-node processing. *IEEE Sensors Journal*, *16*(10), 3972–
749 3980. <https://doi.org/10.1109/JSEN.2016.2530019>
- 750 Jensen, U., Prade, F., & Eskofier, B. M. (2013). Classification of kinematic swimming data with
751 emphasis on resource consumption. In *2013 IEEE International Conference on Body Sensor*
752 *Networks, BSN 2013*. <https://doi.org/10.1109/BSN.2013.6575501>
- 753 Jensen, U., Schmidt, M., Hennig, M., Dassler, F. A., Jaitner, T., & Eskofier, B. M. (2015). An
754 IMU-based mobile system for golf putt analysis. *Sports Engineering*, *18*(2), 123–133.
755 <https://doi.org/10.1007/s12283-015-0171-9>
- 756 Ji, S., Yang, M., Yu, K., & Xu, W. (2013). 3D convolutional neural networks for human action
757 recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 221–
758 231. <https://doi.org/10.1109/TPAMI.2012.59>
- 759 Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Multi-sensor Golf Swing Classification
760 Using Deep CNN. *Procedia Computer Science*, *129*, 59–65.
761 <https://doi.org/10.1016/j.procs.2018.03.046>
- 762 Kapela, R., Świetlicka, A., Rybarczyk, A., Kolanowski, K., & O'Connor, N. E. (2015). Real-time
763 event classification in field sport videos. *Signal Processing: Image Communication*, *35*, 35–
764 45. <https://doi.org/10.1016/j.image.2015.04.005>
- 765 Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014a). Large-
766 scale video classification with convolutional neural networks. *Computer Vision and Pattern*
767 *Recognition (CVPR), 2014 IEEE Conference On*, 1725–1732.
768 <https://doi.org/10.1109/CVPR.2014.223>
- 769 Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014b). Large-
770 scale video classification with convolutional neural networks. Retrieved December 18, 2017,
771 from <http://cs.stanford.edu/people/karpathy/deepvideo/>
- 772 Kasiri-Bidhendi, S., Fookes, C., Morgan, S., Martin, D. T., & Sridharan, S. (2015). Combat sports
773 analytics: Boxing punch classification using overhead depth imagery. In *2015 IEEE*
774 *International Conference on Image Processing (ICIP)* (pp. 4545–4549). Quebec City,
775 Canada: IEEE. <https://doi.org/10.1109/ICIP.2015.7351667>
- 776 Kasiri, S., Fookes, C., Sridharan, S., & Morgan, S. (2017). Fine-grained action recognition of
777 boxing punches from depth imagery. *Computer Vision and Image Understanding*, *159*, 143–
778 153. <https://doi.org/10.1016/j.cviu.2017.04.007>
- 779 Kautz, T. (2017). Acquisition, filtering and analysis of positional and inertial data in sports. *FAU*
780 *Studies in Computer Science*, *2*.
- 781 Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity
782 recognition in beach volleyball using a deep convolutional neural network. *Data Mining and*
783 *Knowledge Discovery*, 1–28. <https://doi.org/10.1007/s10618-017-0495-0>
- 784 Ke, S. R., Thuc, H., Lee, Y. J., Hwang, J. N., Yoo, J. H., & Choi, K. H. (2013). A review on video-
785 based human activity recognition. *Computers*, *2*, 88–131.
786 <https://doi.org/10.3390/computers2020088>
- 787 Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions
788 in elite level rugby union using a wearable sensing device. *Sports Engineering*, *15*(2), 81–92.
789 Retrieved from [https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-](https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5)
790 [0088-5](https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5)
- 791 Kobsar, D., Osis, S. T., Hettinga, B. A., & Ferber, R. (2014). Classification accuracy of a single tri-
792 axial accelerometer for training background and experience level in runners. *Journal of*
793 *Biomechanics*, *47*(10), 2508–2511. <https://doi.org/10.1016/j.jbiomech.2014.04.017>
- 794 Kos, M., & Kramberger, I. (2017). A wearable device and system for movement and biometric data

- 795 Acquisition for sports applications. *IEEE Access*, 1–1.
796 <https://doi.org/10.1109/ACCESS.2017.2675538>
- 797 Kos, M., Ženko, J., Vlaj, D., & Kramberger, I. (2016). Tennis stroke detection and classification
798 using miniature wearable IMU device. In *International Conference on Systems, Signals, and*
799 *Image Processing*. <https://doi.org/10.1109/IWSSIP.2016.7502764>
- 800 Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of
801 classification techniques. *Informatica*, 31, 501–520. <https://doi.org/10.1115/1.1559160>
- 802 Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep
803 convolutional neural networks. *Advances In Neural Information Processing Systems*, 1097–
804 1105. <https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007>
- 805 Lai, D. T. H., Hetchl, M., Wei, X., Ball, K., & McLaughlin, P. (2011). On the difference in swing
806 arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors.
807 *Procedia Engineering*, 13, 219–225. <https://doi.org/10.1016/j.proeng.2011.05.076>
- 808 LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to
809 document recognition. *IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- 810 LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998). Efficient backprop. In *Neural Networks:*
811 *Tricks of the Trade* (Vol. 1524, pp. 9–50).
- 812 LeCun, Y., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, 521(7553), 436–444.
813 <https://doi.org/10.1038/nature14539>
- 814 Li, J., Tian, Q., Zhang, G., Zheng, F., Lv, C., & Wang, J. (2018). Research on hybrid information
815 recognition algorithm and quality of golf swing. *Computers and Electrical Engineering*, 1–
816 13. <https://doi.org/10.1016/j.compeleceng.2018.02.013>
- 817 Liao, W. H., Liao, Z. X., & Liu, M. J. (2003). Swimming style classification from video sequences.
818 In Kinmen (Ed.), *16th IPPR Conference on Computer Vision, Graphics and Image*
819 *Processing* (pp. 226–233). ROC.
- 820 Lu, W. L., Okuma, K., & Little, J. J. (2009). Tracking and recognizing actions of multiple hockey
821 players using the boosted particle filter. *Image and Vision Computing*, 27(1–2), 189–205.
822 <https://doi.org/10.1016/j.imavis.2008.02.008>
- 823 Magalhaes, F. A. de, Vannozzi, G., Gatta, G., & Fantozzi, S. (2015). Wearable inertial sensors in
824 swimming motion analysis: A systematic review. *Journal of Sports Sciences*, 33(7), 732–745.
825 <https://doi.org/10.1080/02640414.2014.962574>
- 826 Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical
827 activity from on-body accelerometers. *Sensors*, 10(2), 1154–1175.
828 <https://doi.org/10.3390/s100201154>
- 829 McNamara, D. J., Gabbett, T., Blanch, P., & Kelly, L. (2017). The relationship between wearable
830 microtechnology device variables and cricket fast bowling intensity. *International Journal of*
831 *Sports Physiology and Performance*, 1–20. <https://doi.org/https://doi.org/10.1123/ijsp.2016-0540>
- 832 McNamara, D. J., Gabbett, T., Chapman, P., Naughton, G., & Farhart, P. (2015). The validity of
833 microsensors to automatically detect bowling events and counts in cricket fast bowlers.
834 *International Journal of Sports Physiology and Performance*, 10(1), 71–75.
835 <https://doi.org/10.1123/ijsp.2014-0062>
- 836 Minnen, D., Westeyn, T. L., Starner, T., Ward, J. a., & Lukowicz, P. (2006). Performance metrics
837 and evaluation issues for continuous activity recognition. In *Proc. Int. Workshop on*
838 *Performance Metrics for Intelligent Systems* (pp. 141–148).
839 <https://doi.org/10.1145/1889681.1889687>
- 840 Mitchell, E., Monaghan, D., & O'Connor, N. E. (2013). Classification of sporting activities using
841 smartphone accelerometers. *Sensors (Basel, Switzerland)*, 13(4), 5317–5337.
842 <https://doi.org/10.3390/s130405317>
- 843 Mohammed, M., Khan, M., & Bashier, E. (2016). *Machine Learning: Algorithms and Applications*.
844 Milton: CRC Press.
- 845 Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting
846 items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7),
847 e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- 848 Montoliu, R., Martín-Félez, R., Torres-Sospedra, O., & Martínez-Usó, A. (2015). Team activity
849 recognition in Association football using a bag-of-words-based method. *Human Movement*
850 *Science*, 41, 165–178. <https://doi.org/10.1016/j.humov.2015.03.007>
- 851 Mooney, R., Corley, G., Godfrey, A., Quinlan, L. R., & ÓLaighin, G. (2015). Inertial sensor
852 technology for elite swimming performance analysis: A systematic review. *Sensors*, 16(1),
853

- 854 18. <https://doi.org/10.3390/s16010018>
- 855 Nibali, A., He, Z., Morgan, S., & Greenwood, D. (2017). Extraction and classification of diving
856 clips from continuous video footage. *ArXiv, pre-print*. Retrieved from
857 <https://arxiv.org/pdf/1705.09003.pdf>
- 858 O'Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017a). Classification
859 of lunge biomechanics with multiple and individual inertial measurement units. *Sports*
860 *Biomechanics*, 16(3), 342–360. <https://doi.org/10.1080/14763141.2017.1314544>
- 861 O'Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017b). Technology in
862 strength and conditioning tracking lower-limb exercises with wearable sensors. *Journal of*
863 *Strength and Conditioning Research*, 31(6), 1726–1736.
- 864 O'Reilly, M., Caulfield, B., Ward, T., Johnston, W., & Doherty, C. (2018). Wearable Inertial
865 Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review.
866 *Sports Medicine*. <https://doi.org/10.1007/s40279-018-0878-4>
- 867 O'Reilly, M., Whelan, D., Chaniavidis, C., Friel, N., Delahunt, E., Ward, T., & Caulfield, B.
868 (2015). Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE*
869 *12th International Conference on Wearable and Implantable Body Sensor Networks*. IEEE.
870 <https://doi.org/10.1109/BSN.2015.7299380>
- 871 O'Reilly, M., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017). Classification of
872 deadlift biomechanics with wearable inertial measurement units. *Journal of Biomechanics*,
873 58, 155–161. <https://doi.org/10.1080/14763141.2017.1314544>
- 874 Ó Conaire, C., Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., & Buckley, J. (2010).
875 Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings*
876 *of the first ACM international workshop on Analysis and retrieval of tracked events and*
877 *motion in imagery streams* (pp. 51–56). ACM. <https://doi.org/10.1145/1877868.1877882>
- 878 Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength
879 training exercises with wearable sensors. *Journal of Biomedical Informatics*, 58, 145–155.
880 <https://doi.org/10.1016/j.jbi.2015.09.020>
- 881 Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in
882 ubiquitous computing. *International Joint Conference on Artificial Intelligence (IJCAI)*,
883 1729.
- 884 Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision*
885 *Computing*, 28(6), 976–990. <https://doi.org/10.1016/j.imavis.2009.11.014>
- 886 Preece, S. J., Goulermas, J. Y., Kenney, L., & Howard, D. (2009). A comparison of feature
887 extraction methods for the classification of dynamic activities from accelerometer data. *IEEE*
888 *Transactions on Biomedical Engineering*, 56(3), 871–879.
889 <https://doi.org/10.1109/TBME.2008.2006190>
- 890 Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009).
891 Activity identification using body-mounted sensors: A review of classification techniques.
892 *Physiological Measurement*, 30(4), R1–R33. <https://doi.org/10.1088/0967-3334/30/4/R01>
- 893 Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine*
894 *Learning*, 42(3), 203–231. <https://doi.org/10.1023/A:1007601015854>
- 895 Qaisar, S., Imtiaz, S., Glazier, P., Farooq, F., Jamal, A., Iqbal, W., & Lee, S. (2013). A method for
896 cricket bowling action classification and analysis using a system of inertial sensors. In
897 *International Conference on Computational Science and its Applications* (pp. 396–412).
898 Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-39649-6>
- 899 Ramanathan, V., Huang, J., Abu-El-Hajja, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2015).
900 Detecting events and key actors in multi-person videos.
901 <https://doi.org/10.1109/CVPR.2016.332>
- 902 Rassem, A., El-Beltagy, M., & Saleh, M. (2017). Cross-country skiing gears classification using
903 deep learning. *ArXiv Preprint ArXiv:1706.08924*. Retrieved from
904 <https://arxiv.org/pdf/1706.08924v1.pdf>
- 905 Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor
906 data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health*
907 *Informatics*, 21(1), 1–1. <https://doi.org/10.1109/JBHI.2016.2633287>
- 908 Reily, B., Zhang, H., & Hoff, W. (2017). Real-time gymnast detection and performance analysis
909 with a portable 3D camera. *Computer Vision and Image Understanding*, 159, 154–163.
910 <https://doi.org/10.1016/j.cviu.2016.11.006>
- 911 Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic
912 classification of sub-techniques in classical cross-country skiing using a machine learning

- 913 algorithm on micro-sensor data. *Sensors (Switzerland)*, 18(1), 75.
914 <https://doi.org/10.3390/s18010075>
- 915 Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using
916 deep learning neural networks. *Expert Systems with Applications*, 59, 235–244.
917 <https://doi.org/10.1016/j.eswa.2016.04.032>
- 918 Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events
919 in soccer videos. *Applied Artificial Intelligence*, 27(7), 656–667.
920 <https://doi.org/10.1080/08839514.2013.787779>
- 921 Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling
922 action in the game of cricket. *Data Mining and Knowledge Discovery*, 31(6), 1706–1734.
923 <https://doi.org/10.1007/s10618-017-0511-4>
- 924 Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B. M. (2015). Inertial
925 sensor-based approach for shot/ pass classification during a soccer match. In *Proc. 21st ACM*
926 *KDD Workshop on Large-Scale Sports Analytics* (pp. 1–4). Sydney, Australia.
- 927 Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving
928 classification performance when training data is skewed. In *9th International Conference on*
929 *Pattern Recognition* (pp. 1–4). <https://doi.org/10.1109/ICPR.2008.4761297>
- 930 Shah, H., Chokalingam, P., Paluri, B., & Pradeep, N. (2007). Automated stroke classification in
931 tennis. *Image Analysis and Recognition*, 1128–1137.
- 932 Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to*
933 *algorithms*. New York, USA: Cambridge University Press.
- 934 Sharma, M., Srivastava, R., Anand, A., Prakash, D., & Kaligounder, L. (2017). Wearable motion
935 sensor based phasic analysis of tennis serve for performance feedback. In *2017 IEEE*
936 *International Conference on Acoustics, Speech and Signal Processing* (pp. 5945–5949). New
937 Orleans, LA: IEEE.
- 938 Sprager, S., & Juric, M. B. (2015). *Inertial sensor-based gait recognition: A review. Sensors*
939 *(Switzerland)* (Vol. 15). <https://doi.org/10.3390/s150922089>
- 940 Srivastava, R., Patwari, A., Kumar, S., Mishra, G., Kaligounder, L., & Sinha, P. (2015). Efficient
941 characterization of tennis shots and game analysis using wearable sensors data. In *2015 IEEE*
942 *Sensors- Proceedings* (pp. 1–4). Busan. <https://doi.org/10.1109/ICSENS.2015.7370311>
- 943 Stein, M., Janetzko, H., Lamprecht, A., Breikreutz, T., Zimmermann, P., Goldlücke, B., ... Keim,
944 D. A. (2018). Bring it to the pitch: combining video and movement data to enhance team
945 sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 13–22.
946 <https://doi.org/10.1109/TVCG.2017.2745181>
- 947 Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks:
948 A tutorial and survey. *IEEE*, 105(2), 2295–2329. Retrieved from
949 <http://arxiv.org/abs/1703.09039>
- 950 Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports:
951 Current applications and research topics. *Computer Vision and Image Understanding*, 159, 3–
952 18. <https://doi.org/10.1016/j.cviu.2017.04.011>
- 953 Titterton, D. H., & Weston, J. L. (2009). *Strapdown inertial navigation technology* (2nd ed.).
954 Reston, VA: AIAA.
- 955 Tora, M. R., Chen, J., & Little, J. J. (2017). Classification of puck possession events in ice hockey.
956 In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*
957 *Workshops* (pp. 147–154). <https://doi.org/10.1109/CVPRW.2017.24>
- 958 Victor, B., He, Z., Morgan, S., & Miniutti, D. (2017). Continuous video to simple signals for
959 swimming stroke detection with convolutional neural networks. *ArXiv Preprint*
960 *ArXiv:1705.09894*. <https://doi.org/10.1111/j.1467-8330.1974.tb00606.x>
- 961 Wagner, D., Kalischewski, K., Velten, J., & Kummert, A. (2017). Activity recognition using
962 inertial sensors and a 2-D convolutional neural network. In *IEEE (Ed.), 2017 10th*
963 *International Workshop on Multidimensional (nD) Systems (nDS)* (pp. 1–6).
964 <https://doi.org/10.1109/NDS.2017.8070615>
- 965 Wagner, J. F. (2018). About motion measurement in sports based on gyroscopes and
966 accelerometers - an engineering point of view. *Gyroscopy and Navigation*, 9(1), 1–18.
967 <https://doi.org/10.1134/S2075108718010091>
- 968 Ward, J. A., Lukowicz, P., & Gellersen, H.-W. (2011). Performance metrics for activity
969 recognition. In *ACM Trans. on Intelligent Systems and Technology* (Vol. 2, pp. 111–132).
- 970 Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring hitting load in tennis using
971 inertial sensors and machine learning. *International Journal of Sports Physiology and*

- 972 *Performance*, 1–20. <https://doi.org/https://doi.org/10.1123/ijssp.2016-0683>
- 973 Wixted, A., Billing, D. C., & James, D. A. (2010). Validation of trunk mounted inertial sensors for
974 analysing running biomechanics under field conditions, using synchronously collected foot
975 contact data. *Sports Engineering*, 12(4), 207–212. <https://doi.org/10.1007/s12283-010-0043-2>
- 976 Wixted, A., Portus, M., Spratford, W., & James, D. A. (2011). Detection of throwing in cricket
977 using wearable sensors. *Sports Technology*, 4(3–4), 134–140.
978 <https://doi.org/10.1080/19346182.2012.725409>
- 979 Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural*
980 *Computation*, 8(7), 1341–1390. <https://doi.org/10.1162/neco.1996.8.7.1391>
- 981 Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimisation. *IEEE*
982 *Transactions on Evolutionary Computation*, 1(1), 67–82.
983 <https://doi.org/10.1023/A:1021251113462>
- 984 Wundersitz, D. W., Gastin, P. B., Richter, C., Robertson, S., & Netto, K. J. (2015). Validity of a
985 trunk-mounted accelerometer to assess peak accelerations during walking, jogging and
986 running. *European Journal of Sport Science*, 15(5), 382–390.
987 <https://doi.org/10.1080/17461391.2014.955131>
- 988 Wundersitz, D. W., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of
989 a trunk-mounted accelerometer to measure peak impacts during team sport movements.
990 *International Journal of Sports Medicine*, 36(9), 742–746. <https://doi.org/10.1055/s-0035-1547265>
- 991 Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015).
992 Classification of team sport activities using a single wearable tracking device. *Journal of*
993 *Biomechanics*, 48(15), 3975–3981. <https://doi.org/10.1016/j.jbiomech.2015.09.015>
- 994 Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for
995 physical activity monitoring. *Sensors*, 10(8), 7772–7788. <https://doi.org/10.3390/s100807772>
- 996 Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Shonali, K. (2015). Deep convolutional neural
997 networks on multichannel time series for human activity recognition. In *Proceedings of the*
998 *24th International Conference on Artificial Intelligence* (pp. 3995–4001).
- 1000 Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object
1001 interaction activities. In *Computer Vision and Pattern Recognition* (pp. 17–24). IEEE.
- 1002 Young, C., & Reinkensmeyer, D. J. (2014). Judging complex movement performances for
1003 excellence: a principal components analysis-based technique applied to competitive diving.
1004 *Human Movement Science*, 36, 107–122. <https://doi.org/10.1016/j.humov.2014.05.009>
- 1005 Yu, G., Jang, Y. J., Kim, J., Kim, J. H., Kim, H. Y., Kim, K., & Panday, S. B. (2016). Potential of
1006 IMU sensors in performance analysis of professional alpine skiers. *Sensors (Switzerland)*,
1007 16(4), 1–21. <https://doi.org/10.3390/s16040463>
- 1008 Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human Activity Recognition with Inertial
1009 Sensors Using a Deep Learning Approach. *Proc. of IEEE Sensors 2016*, (1), 1–3.
1010 <https://doi.org/10.1109/ICSENS.2016.7808590>
- 1011 Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014).
1012 Convolutional neural networks for human activity recognition using mobile sensors. In
1013 *Proceedings of the 6th International Conference on Mobile Computing, Applications and*
1014 *Services* (pp. 197–205). <https://doi.org/10.4108/icst.mobicase.2014.257786>
- 1015 Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity
1016 recognition using vision-based method. *Journal of Healthcare Engineering*, 2017, 1–31.
1017 <https://doi.org/10.1155/2017/3090343>
- 1018 Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-
1019 channels deep convolutional neural networks. In *International Conference on Web-Age*
1020 *Information Management* (pp. 298–310). Springer. https://doi.org/10.1007/978-3-319-08010-9_33
- 1021 Zhu, G., Xu, C., Gao, W., & Huang, Q. (2006). Action recognition in broadcast tennis video.
1022 *Computer Vision in Human-Computer Interaction*, 89–98.
1023 https://doi.org/10.1007/11754336_9
- 1024 Ziaefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review.
1025 *Pattern Recognition*, 48(8), 2329–2345. <https://doi.org/10.1016/j.patcog.2015.03.006>
- 1026
1027

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65