

Concurrent validity of computer-vision artificial intelligence player tracking software using broadcast footage

CRANG, Zachary L <<http://orcid.org/0009-0001-4915-4022>>, JOHNSTON, Rich D, MILLS, Katie, BILLINGHAM, Johsan, ROBERTSON, Sam, COLE, Michael H <<http://orcid.org/0000-0003-1817-7528>>, WEAKLEY, Jonathon, HEWITT, Adam and DUTHIE, Grant M

Available from Sheffield Hallam University Research Archive (SHURA) at:
<https://shura.shu.ac.uk/37394/>

This document is the Published Version [VoR]

Citation:

CRANG, Zachary L, JOHNSTON, Rich D, MILLS, Katie, BILLINGHAM, Johsan, ROBERTSON, Sam, COLE, Michael H, WEAKLEY, Jonathon, HEWITT, Adam and DUTHIE, Grant M (2026). Concurrent validity of computer-vision artificial intelligence player tracking software using broadcast footage. *Journal of Sports Analytics*, 12. [Article]

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Concurrent validity of computer-vision artificial intelligence player tracking software using broadcast footage

Journal of Sports Analytics
Vol. 12(0): 1–13
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/22150218261445834
journals.sagepub.com/home/san



Zachary L Crang¹ , Rich D Johnston^{1,2,3}, Katie L Mills^{4,5} ,
Johsan Billingham⁴, Sam Robertson⁶, Michael H Cole¹ ,
Jonathon Weakley^{1,2,3}, Adam Hewitt¹ and Grant M Duthie^{7,8}

Abstract

This study aimed to: (1) quantify the accuracy of commercially available computer-vision and artificial intelligence (AI) player tracking software to measure player position, speed and distance covered using broadcast footage and (2) determine the impact of camera feed and resolution on accuracy. Data were obtained from one match at the 2022 Qatar Fédération Internationale de Football Association (FIFA) World Cup. Tactical, programme and camera 1 feeds were used. Three commercial tracking providers that use computer-vision and AI participated. Providers analysed instantaneous position (x, y co-ordinates) and speed ($\text{m}\cdot\text{s}^{-1}$) of each player. Their data were compared with a high-definition multi-camera tracking system (TRACAB Gen 5). Root mean square error (RMSE) and mean bias were calculated. Position RMSE ranged from 1.68 to 16.39 m, while speed RMSE ranged from 0.34 to 2.38 $\text{m}\cdot\text{s}^{-1}$. Total distance mean bias ranged from -1745 m (-21.8%) to 1945 m (24.3%) across providers. Computer-vision and AI player tracking software offer the best accuracy when players are detected by the software. Providers should use a tactical feed when tracking position and speed, which will maximise player detection, improving accuracy. Both 720p and 1080p resolutions are suitable, assuming appropriate computer-vision and AI models are implemented.

Keywords

tracking technology, monitoring, accuracy, team sports

Introduction

The quantification and interpretation of a player's match or training activities, often termed external load, is widespread in team sports. It is common across various sports to report both aggregated measures (e.g., total distance) as well as discrete phases of play such as the mean peak speed over a 5-min period (Johnston et al., 2019, 2020; Thoseby et al., 2023; Modric et al., 2020; Whitehead et al., 2019). This match activity data is used to inform specific training prescription, while training loads are monitored live and retrospectively to ensure the intended training outcomes are achieved. Beyond this, tracking data can also be used for tactical analysis, providing insights into spatial organisation, collective movement patterns and tactical behaviours during match play (Goes et al., 2021). The quantification of match-demands therefore needs to be performed with accurate tracking systems that can also capture

¹School of Behavioural and Health Sciences, Australian Catholic University, Brisbane, QLD, Australia

²Sports Performance, Recovery, Injury and New Technologies (SPRINT) Research Centre, Australian Catholic University, Brisbane, QLD, Australia

³Carnegie Applied Rugby Research (CARR) Centre, School of Sport, Leeds Beckett University, Leeds, UK

⁴Fédération Internationale de Football Association (FIFA), Zurich, Switzerland

⁵Sports Engineering Research Group, Sheffield Hallam University, UK

⁶Institute for Health & Sport, Victoria University, Melbourne, Australia

⁷Global Sport and Movement Collaborative, University of Newcastle, New South Wales, Australia

⁸School of Biomedical Sciences and Pharmacy, College of Health, Medicine, and Wellbeing, University of Newcastle, Newcastle, Australia

Corresponding author:

Zachary Crang, School of Behavioural and Health Sciences, Australian Catholic University, Brisbane, QLD, Australia.
Email: zachary.crang@myacu.edu.au



enough of the population to provide representative data (West et al., 2021).

Within professional sport, both optical tracking and wearable microtechnology devices are widely used to measure player activities during matches. Optical tracking has developed from early manual notational analysis (Jaques and Pavia, 1974; Nettleton and Sandstrom, 1963) to semi-automated vision-based tracking systems (Barris and Button, 2008; Duthie et al., 2003), before automated systems that use computer-based image processing techniques evolved (Figueroa et al., 2006; Iwase and Saito, 2003, 2004). Whilst these automated systems can provide a high level of accuracy (Linke et al., 2020), they are not entirely autonomous and often require human intervention – for example, the manual tagging or verification of players when tracking errors occur (Iwase and Saito, 2003, 2004). Further, optical tracking systems require several fixed or specific camera angles to operate, meaning they are generally only suitable for stadiums with appropriate infrastructure. Wearable microtechnology, which includes global navigation satellite system (GNSS) and local positioning systems (LPS) can circumvent some of the issues with optical tracking. These GNSS devices, allow practitioners to easily quantify the locomotive match-demands of players (Johnston et al., 2020). Whilst their portability is a strength, the requirement of players to wear the device in a vest or within the playing jersey may impact their use due to regulations of the sport or player preferences.

With developments in computer-vision and artificial intelligence (AI) software, players can be tracked with extremely limited human intervention, using video footage of the match. To collect the footage, single (e.g., wide-angle lens) cameras (Hurault et al., 2020; Scott et al., 2022; Stein et al., 2017; Thinh et al., 2019) placed on the half-way line, multiple fixed cameras (Xu et al., 2004), and broadcast acquired footage have been used (Mazzeo et al., 2008; Naik and Hashmi, 2021; Stein et al., 2017). Factors that may influence the ability of computer-vision software to accurately detect and track players include occlusion (Gabriel et al., 2003), misidentification of players (Xu et al., 2004), and video resolution (Thinh et al., 2019). Occlusions can occur where players are gathered closely (e.g., set pieces), misidentification may result when players are similar in appearance (e.g., body shape, boot colour). Further, algorithms using video files with better resolution (2.5k vs. 1080p) were generally more accurate at detecting and tracking players (Thinh et al., 2019). Unlike automated fixed camera systems, broadcast cameras do not maintain full-field coverage, meaning players can frequently move outside the frame for extended periods. Consequently, tracking systems must infer the un-detected players' locations during these gaps through imputation and interpolation. Albeit, given there is no equipment or set-up required by teams when

using broadcast footage for player tracking compared to the aforementioned methods, it emerges as the most practical approach. This is particularly advantageous in amateur football and scouting contexts, where teams often lack access to dedicated tracking technologies. However, the agreement between the outcomes derived from broadcast footage and the other previously established systems is not fully understood.

To the authors' knowledge, no current study has investigated the accuracy of computer-vision and AI software that use broadcast footage to track players position (x, y co-ordinates) and speed. Given that there is likely to be a wide range of proprietary methods that the various providers use to collect, process, and generate the data, there may be substantial variability in data accuracy. Consequently, the aim of this study was to: (1) quantify the accuracy of commercially available computer-vision and AI player tracking software to measure player position, speed and distance covered using broadcast footage; (2) determine the impact of camera feed and video resolution on accuracy. It was hypothesised that accuracy would be greater for higher resolution camera feeds that had the players in the camera frame for a greater proportion of time, therefore requiring less imputation (i.e., 1080p tactical feed).

Methods

Data were collected during a single group stage match of the 2022 Qatar Fédération Internationale de Football Association (FIFA) World Cup tournament.

The match was filmed by the television broadcasters using multiple fixed video cameras positioned around the field of play at 50 frames per second (fps) and stored at 25 fps. The tactical, programme and camera 1 video feeds were obtained from the FIFA data platform in MPEG-4 Part-14 (.mp4) file format. The tactical feed (Figure 1C) is a wide-view angle of the pitch, captured from a fixed camera positioned on the top tier of the grandstand, at the half-way line. The purpose of this camera is to maintain 20 outfield players in shot. The programme feed (Figure 1A) comprises multiple broadcast camera angles (e.g., tactical and end-on views) and represents the footage televised to the public, including on-screen graphics and replays. As a result of the frequent use of different camera angles, the feed regularly switches perspectives and exhibits varying levels of zoom. The camera 1 feed (Figure 1B) captures a slightly tighter field of view compared to the tactical feed. It makes up the majority of the programme feed, except it does not have any graphical overlay (e.g., scoreboard), replays or cut to different camera angles.

Figure 1 shows an example of the programme (Figure 1A), camera 1 (Figure 1B) and tactical feeds (Figure 1C), at the same moment in time. These .mp4

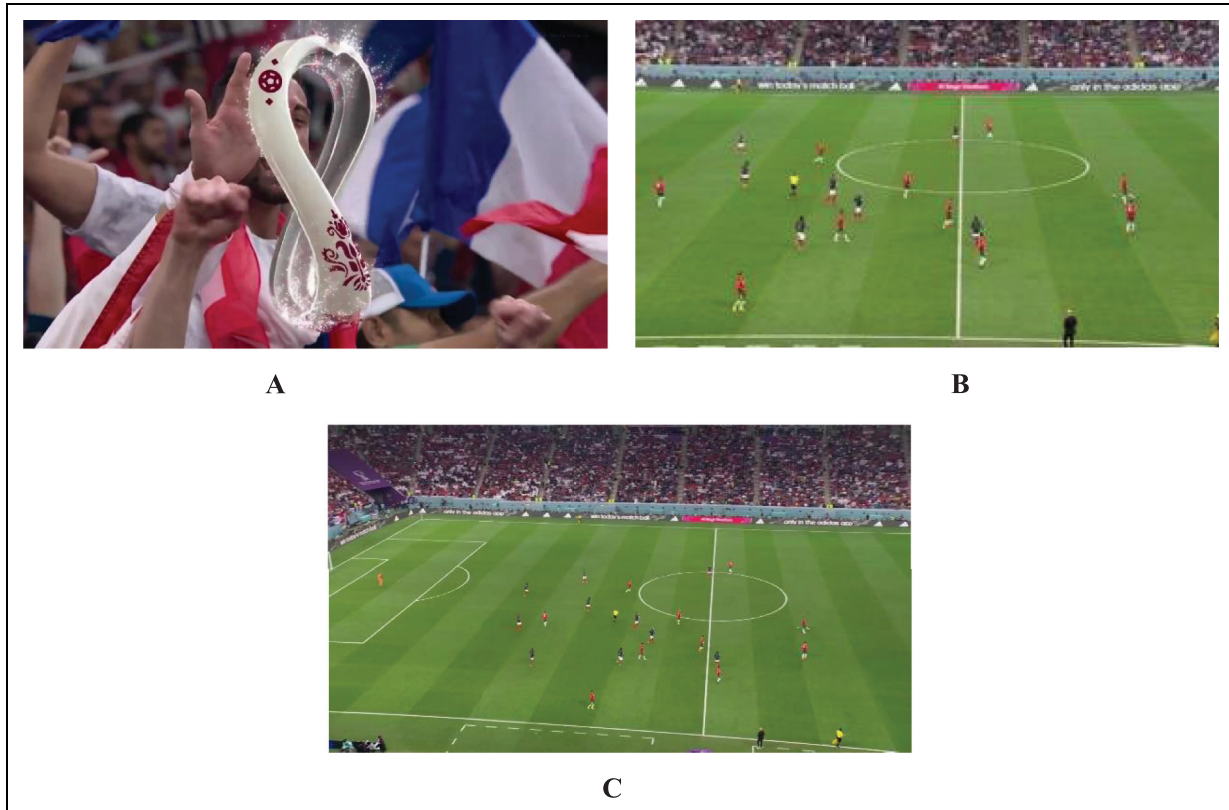


Figure 1. Different camera feeds at the same moment in time. (A) Programme feed, (B) Camera 1 feed, and (C) Tactical feed.

files were downloaded in two different video resolutions; 720p and 1080p. As such, a total of six .mp4 files were provided to the tracking providers to run their analyses.

Only players who were on the field of play, including those who entered as substitutes, were tracked for analysis. Given their differing activity profiles to other positions as well as poor visibility (i.e., outside of the camera frame), goalkeepers were removed from the analysis, resulting in a total of 27 players (Clemente et al., 2013). The match consisted of a 1st half (48 min) and second half (50 min), with no extra time.

Commercially available player tracking providers that use computer-vision and AI to estimate position and speed data were invited by FIFA to take part in this study, with three providers volunteering. These providers specialise in player-tracking across football, basketball, and other team sports, and are all certified under FIFA's Electronic Performance Tracking System (EPTS) quality programme. All providers were founded in 2015 to 2016. A combination of undisclosed computer-vision and AI techniques were used to track the overall position (x , y co-ordinates) and speed ($\text{m}\cdot\text{s}^{-1}$) of each player at 25 fps. Upon completion, the providers delivered the tracking data in comma-separated value (.csv) files in their raw sampling frequency, where each row provided an observation of player speed and position at 25 Hz. The datasets provided

also included detail regarding whether the outcomes were derived when a player was detected or undetected by the provider's software. An example of when the software may not detect the player is when they are positioned outside of the camera's field of view or when a player is obscured by another player. The software recognises when a player is on or off the field. Therefore, when processing the image, if a player is expected to be present but is not detected by the software, the embedded AI model is used to estimate the player's position and speed (Omidshafiei et al., 2022). In this study, the term 'detected' or 'undetected' refers to whether the player was identified by the computer-vision software, while 'visible' indicates that the player was within the camera's field of view.

To establish the concurrent validity, providers were compared against a high-definition multi-camera optical tracking system (TRACAB Gen 5, ChyronHego, New York, USA) capturing data at 25 Hz. TRACAB is a fixed camera system using 12 cameras elevated within the stadium infrastructure. This system has strong accuracy for measures of position ($\text{RMSE}=0.08$ m) and speed ($\text{RMSE}=0.08$ $\text{m}\cdot\text{s}^{-1}$) compared to 3D motion capture (Linke et al., 2020).

For consistency, the speed from each provider and the TRACAB system were filtered using a 4th order 1 Hz low-pass Butterworth filter (Delves et al., 2022).

Individual player tracking data from each provider for each video resolution were temporally aligned by phase shifting the players' speeds within the TRACAB data to establish the smallest RMSE. The most common phase shift (mode) was then applied to all files. The two datasets were then spatially aligned by rotating the providers' tracking data throughout 360 degrees in 0.1-degree increments to establish the smallest mean error in XY position. Further, the X and Y co-ordinates were spatially shifted to overlay each other. The most common rotation and shift was found across all files and then applied. The accuracy of the providers to measure overall position and speed was assessed. In addition, the total distance across the entire match was also examined to provide an understanding of how position and speed error influence aggregate data. Total distance was calculated for each player by multiplying their change in speed by change in time.

Statistical analysis

The statistical analyses were performed in RStudio (version 12.1; Posit, Boston, MA) using the R programming language (version 4.3.3, R Foundation for Statistical Computing, Vienna, Austria).

First, to determine the agreement between the different providers and TRACAB, mean bias and 95% the limits of agreement (LOA) were estimated from linear mixed effects (position and speed) and linear models (total distance) built using the `lmerTest::lmer` and `stats::lm` functions (Supplementary Table 1). Separate models were built for speed, position and total distance, for each camera (Programme, Camera 1, Tactical) and video resolution (720p, 1080p). For the speed and total distance analyses, the error (i.e., difference to TRACAB) was specified as the outcome variable, with the corresponding TRACAB measure included as a fixed-effect. In the speed models, Player ID was included as a random effect (Parker et al., 2016). The 95% LOA were derived from these models. Mean bias was estimated using the same model structure but with the fixed-effect removed. For position, mean bias and 95% LOA were obtained from intercept-only models, with positional error used as the outcome variable and Player ID as a random effect. All models were run using three different datasets: one including only data where the players were detected by the software, one with only data where they were undetected, and a combined dataset that included all data. The root mean square error (RMSE) was separately calculated for position and speed to quantify absolute error. In line with the aim of the study to quantify tracking accuracy, RMSE values are reported without comparison to predefined accuracy thresholds. At times, the suitability of the observed accuracy for specific use cases are provided to help provide context; the reader should use their own judgement to determine whether the

magnitude of error is acceptable for their intended application.

$$RMSE = \sqrt{\frac{\sum^N (\text{Manufacturer Speed} - \text{TRACAB Speed})^2}{N}}$$

Where N is the number of observations in the raw data.

Second, to establish the influence of provider, camera feed (e.g., programme vs. tactical) and video resolution (720p vs. 1080p) on the accuracy of speed, position and total distance, linear mixed models were fit using REML (Supplementary Table 1). The RMSE for position and speed, as well as mean bias for total distance were used as the outcome variables, with provider, camera feed and video resolution used as fixed-effects in a three-way interaction; player ID was incorporated as a random intercept term. The main effects from each model were extracted using the `stats::anova` function. Where significant main effects were observed, post hoc tests were performed using the `emmeans::emmeans` function with a Tukey adjustment applied to account for multiple comparisons. Data are presented as mean \pm SD; statistical significance was set at $p < 0.05$. The linear mixed-effects models used in this study assume linear relationships between predictors and outcomes, normally distributed random effects, and residual errors that are independent and homoscedastic conditional on the random effects. These assumptions were assessed via visual inspection of diagnostic plots using the `check_model` function from the `performance` package. Overall, assumptions were reasonably satisfied, although some evidence of heteroscedasticity was observed. To address this, additional models explicitly accounting for the variance structure were fitted; however, this did not alter the results. The inclusion of the random intercept accounts for within-cluster dependence, such that remaining residuals are assumed independent given the random deviations from normality, particularly with moderate to large numbers of clusters (Schielzeth et al., 2020).

Results

Data analysed and detected

The number of datapoints analysed by each provider and percentage within each configuration is outlined in Table 1.

Positional accuracy

For positional accuracy (Table 2), there were significant main effects for provider ($p < 0.001$), camera feed ($p < 0.001$), and video resolution ($p < 0.001$). The configuration that produced the best accuracy was different across providers. The best accuracy for position was seen for Provider 2 using the 720p tactical feed (RMSE = 1.68 m; mean bias

Table 1. The number (% total) of player-frames detected by each provider using the programme (PGM), camera 1 and tactical feeds at the 720p and 1080p resolutions.

		Datapoints (%)					
		720p			1080p		
		PGM	Cam 1	Tactical	PGM	Cam 1	Tactical
Provider 1	Detected	1,236,404 (42%)	1,883,411 (63%)	2,802,310 (94%)	1,236,589 (42%)	1,881,219 (63%)	2,819,394 (95%)
	Undetected	1,732,652 (58%)	1,092,729 (37%)	173,723 (6%)	1,732,467 (58%)	1,094,921 (37%)	156,639 (5%)
	Overall	2,969,056	2,976,140	2,976,033	2,969,056	2,976,140	2,976,033
Provider 2	Detected	1,078,425 (36%)	1,661,976 (56%)	2,647,867 (89%)	1,125,202 (38%)	1,726,834 (58%)	2,680,068 (90%)
	Undetected	1,881,935 (64%)	1,298,482 (44%)	316,005 (11%)	1,835,158 (62%)	1,233,624 (42%)	283,800 (10%)
	Overall	2,963,872	2,960,458	2,963,872	2,963,872	2,960,458	2,963,868
Provider 3	Detected	1,203,392 (42%)	1,845,274 (64%)	2,775,052 (96%)	1,202,991 (42%)	1,846,269 (64%)	2,757,273 (95%)
	Undetected	1,690,595 (58%)	1,048,713 (36%)	119,418 (4%)	1,690,996 (58%)	1,048,285 (36%)	137,197 (5%)
	Overall	2,893,987	2,893,987	2,894,470	2,893,987	2,894,554	2,894,470

PGM: Programme feed.

Table 2. Concurrent validity of player software tracking systems to measure position (m) during football match-play in comparison to an optical tracking system.

		Mean bias (m) ± 95% LOA																	
		RMSE (m)						720p						1080p					
		720p		1080p		Tactical		Cam 1		PGM		Tactical		Cam 1		PGM		Tactical	
Provider 1	Detected	1.14	2.23	1.79	1.07	1.69	1.68	0.53 ± 1.99	0.79 ± 4.09	0.82 ± 3.27	0.54 ± 1.82	0.70 ± 3.01	0.80 ± 2.94						
	Undetected	4.62	9.56	10.82	4.58	9.65	12.23	2.80 ± 7.24	5.29 ± 16.62	5.74 ± 17.09	2.81 ± 7.13	5.14 ± 16.08	6.13 ± 20.66						
	Overall	3.14 ^a	6.29 ^b	3.14	3.10 ^a	6.00 ^b	3.25	1.50 ± 5.42	2.40 ± 11.38	1.17 ± 5.89	1.51 ± 5.32	2.31 ± 10.85	1.11 ± 6.02						
Provider 2	Detected	1.13	0.66	0.47	1.02	0.68	0.44	1.03 ± 0.92	0.60 ± 0.59	0.41 ± 0.47	0.95 ± 0.77	0.59 ± 0.70	0.37 ± 0.47						
	Undetected	11.56	9.27	4.96	9.57	9.08	6.19	7.75 ± 16.58	6.42 ± 13.24	2.95 ± 8.15	6.57 ± 13.72	6.15 ± 13.25	3.59 ± 9.66						
	Overall	9.24 ^{a,b}	6.16 ^b	1.68	7.57 ^b	5.89 ^b	1.96	5.36 ± 14.71	3.14 ± 10.42	0.68 ± 3.03	4.45 ± 12.04	2.91 ± 10.08	0.72 ± 3.61						
Provider 3	Detected	3.62	5.03	16.24	3.05	2.49	9.95	1.48 ± 6.49	2.78 ± 8.15	8.69 ± 26.75	1.37 ± 5.38	0.85 ± 4.58	3.34 ± 17.61						
	Undetected	10.03	10.27	19.38	10.02	9.83	15.03	6.37 ± 15.11	6.53 ± 15.46	11.25 ± 26.16	6.33 ± 15.17	5.96 ± 15.30	7.67 ± 23.49						
	Overall	8.01 ^b	7.37 ^b	16.39 ^c	7.91	6.24	10.25	4.31 ± 13.18	4.11 ± 11.91	8.78 ± 26.70	4.24 ± 13.04	2.66 ± 11.03	3.55 ± 18.05						

RMSE; root mean square error; LOA; limits of agreement; PGM; Programme feed.

^aSignificant difference ($p < 0.05$) to cam 1 feed for the same provider and resolution.

^bSignificant difference ($p < 0.05$) to tactical camera feed for the same provider and resolution.

^cSignificant difference ($p < 0.05$) between camera resolution for the same provider and camera feed.

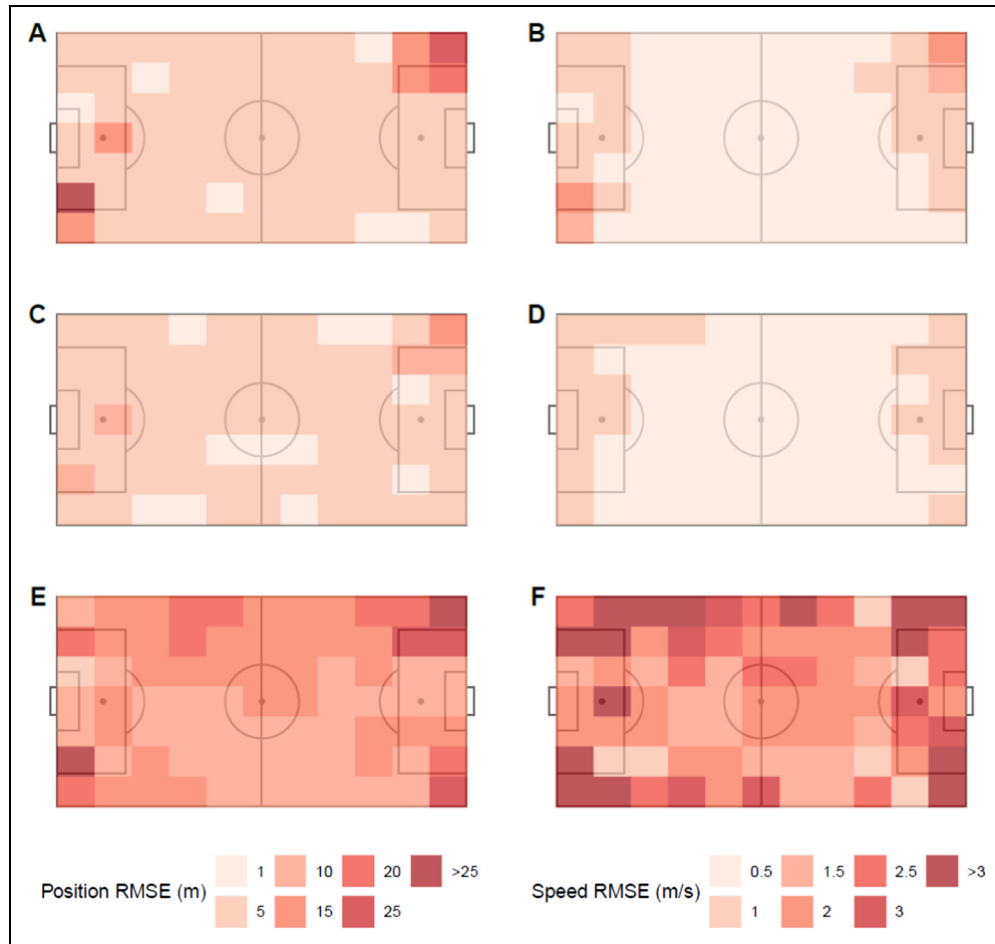


Figure 2. Location of position (A, C, E) and speed (B, D, F) error for providers 1 (A, B), 2 (C, D) and 3 (E, F).

[LOA] = 0.68 m [−2.35 to 3.71 m]). The best for Provider 1 was the 1080p programme feed (RMSE = 3.14 m; mean bias [LOA] = 1.51 m [−3.81 to 6.83 m]), while for Provider 3, the best accuracy was seen for the 1080p camera 1 feed (RMSE = 6.24 m; mean bias [LOA] = 2.66 m [−8.37 to 13.69 m]). The location of the positional error for the 1080p tactical feed across the 3 providers is illustrated in Figure 2(A), (C) and (E). The cumulative proportion of positional error (i.e., difference to criterion) is shown in Figure 3(A), (C) and (E). To aid with comparisons between configurations and providers, the proportion of positional error under 1 m for the programme feed (Figure 3A), was 69.3% for Provider 1, 31.9% for Provider 2, and 24.1% for Provider 3; camera 1 feed (Figure 3C), 63.7% for Provider 1, 60.5% for Provider 2, and 60.7% for Provider 3; for the tactical feed (Figure 3E), 84.1% for Provider 1, 93.7% for Provider 2, and 82.6% for Provider 3.

Across configurations of camera feed and resolution, Provider 1 had the most consistent accuracy for position compared to TRACAB (RMSE = 3.10 to 6.29 m), with significantly better ($p < 0.001$) accuracy for the programme

feed and tactical feed compared to the camera 1 feed. The accuracy of Providers 2 and 3 was more variable across configurations, with several significant differences to TRACAB observed (Table 2). Specifically, for Provider 2, there were larger errors for the programme feed at 720p compared to the camera 1 ($p < 0.001$) and tactical feed ($p < 0.001$). The camera 1 feed was also worse compared to the tactical feed ($p < 0.001$). Similarly, at 1080p, the tactical feed was significantly more accurate compared to the camera 1 feed ($p < 0.001$) and the programme feed ($p < 0.001$). For Provider 3, their tactical feed was significantly worse at 720p compared to the programme ($p < 0.001$) and camera 1 ($p < 0.001$) feeds; this configuration was also significantly worse than the tactical feed at 1080p ($p < 0.001$).

Speed accuracy

For speed accuracy (Table 3), there were significant main effects of provider ($p < 0.001$), camera feed ($p < 0.001$), and video resolution ($p < 0.001$). The configuration that

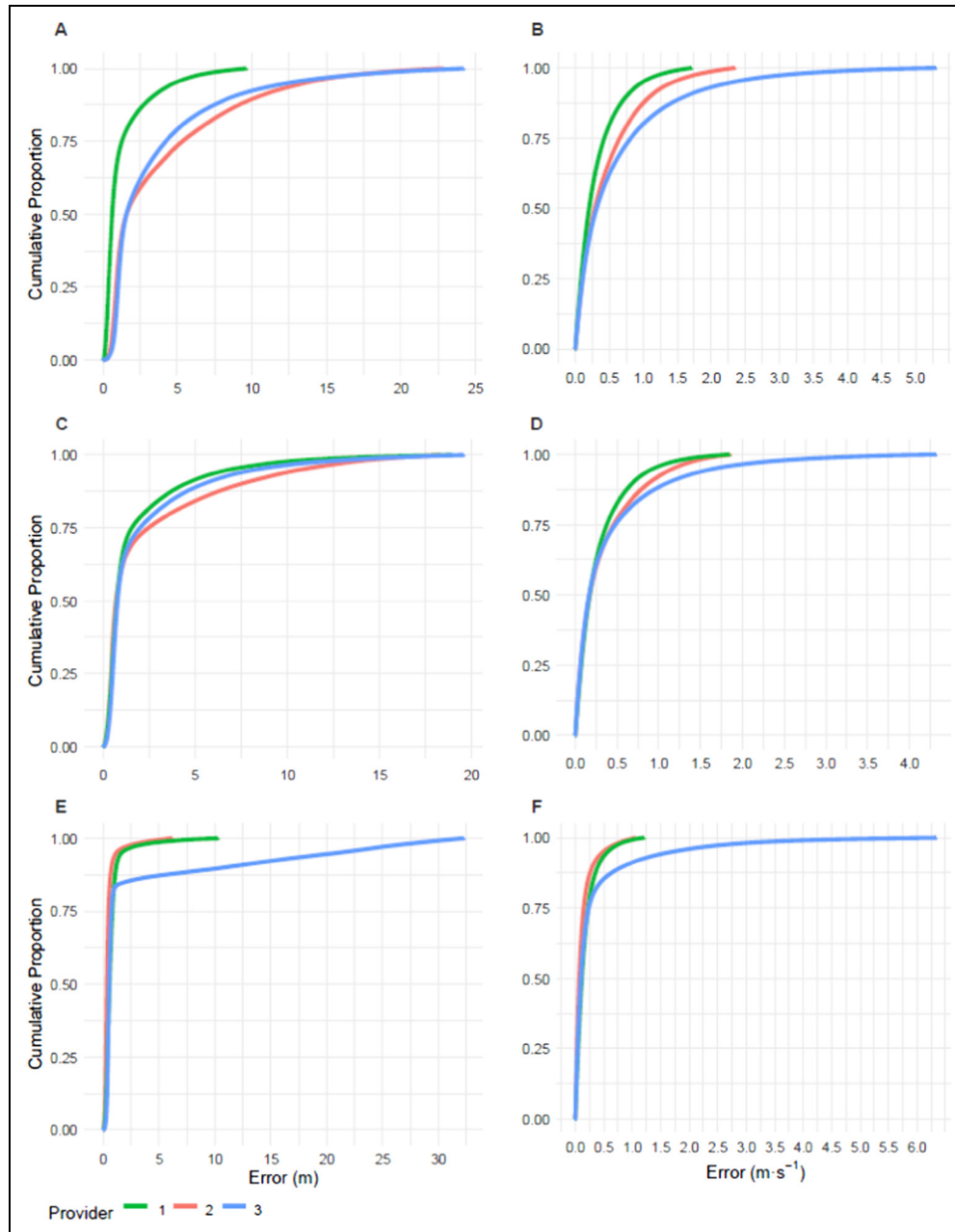


Figure 3. Cumulative proportion of error (i.e., difference to criterion) for position (A, C, D) and speed (B, E, F) for programme (A, B), camera 1 (C, D) and tactical (E, F) 1080p feeds. The y-axis represents the cumulative percentage of observations with an error less than or equal to the corresponding x-axis value.

produced the best accuracy was different across providers. The best accuracy for speed was seen by Provider 2 using the 1080p tactical feed (RMSE = $0.34 \text{ m}\cdot\text{s}^{-1}$; mean bias [LOA] = $0.02 \text{ m}\cdot\text{s}^{-1}$ [-0.63 to $0.67 \text{ m}\cdot\text{s}^{-1}$]). The best for Provider 1 was using the 1080p tactical feed (RMSE = $0.39 \text{ m}\cdot\text{s}^{-1}$; mean bias [LOA] = $-0.06 \text{ m}\cdot\text{s}^{-1}$ [-0.80 to $0.68 \text{ m}\cdot\text{s}^{-1}$]) while for Provider 3, the best was for the 1080p camera 1 feed (RMSE = $1.37 \text{ m}\cdot\text{s}^{-1}$; mean bias [LOA] = $0.26 \text{ m}\cdot\text{s}^{-1}$ [-2.32 to $2.84 \text{ m}\cdot\text{s}^{-1}$]). The location of the speed error for the 1080p tactical feed across the 3

providers is illustrated in Figure 2(B), (D) and (F). The cumulative proportion of speed error (i.e., difference to criterion) is shown in Figure 3(B), (D) and (F). For the programme feed (Figure 3B), the percentage of speed error under $1.0 \text{ m}\cdot\text{s}^{-1}$ was 95.1% for Provider 1, 87.5% for Provider 2, and 80.0% for Provider 3. For the camera 1 feed (Figure 3D), the proportion of speed error under $1.0 \text{ m}\cdot\text{s}^{-1}$ was 95.8% for Provider 1, 92.3% for Provider 2, and 88.4% for Provider 3. For the tactical feed, the percentage of speed error under $1.0 \text{ m}\cdot\text{s}^{-1}$ was 99.3% for

Table 3. Concurrent validity of player software tracking systems to measure speed ($\text{m}\cdot\text{s}^{-1}$) during football match-play in comparison to an optical tracking system.

	Mean bias ($\text{m}\cdot\text{s}^{-1}$) \pm 95% LOA																	
	RMSE ($\text{m}\cdot\text{s}^{-1}$)						720p						1080p					
	720p		1080p		Tactical		PGM		Cam 1		Tactical		PGM		Cam 1		Tactical	
Provider 1	Detected	0.35	0.34	0.35	0.35	0.32	0.32	0.32	0.32	-0.10 \pm 0.65	-0.08 \pm 0.64	-0.08 \pm 0.66	-0.10 \pm 0.65	-0.08 \pm 0.61	-0.07 \pm 0.61			
	Undetected	0.78	0.91	1.10	0.78	0.89	1.02	0.89	1.02	-0.20 \pm 1.34	-0.16 \pm 1.62	-0.06 \pm 1.80	-0.20 \pm 1.34	-0.16 \pm 1.59	0.05 \pm 1.72			
	Overall	0.57	0.61	0.43	0.57	0.60	0.39	0.60	0.39	-0.14 \pm 1.06	-0.11 \pm 1.15	-0.08 \pm 0.81	-0.14 \pm 1.06	-0.11 \pm 1.13	-0.06 \pm 0.74			
Provider 2	Detected	0.39	0.27	0.24	0.46	0.30	0.20	0.30	0.20	-0.04 \pm 0.76	-0.004 \pm 0.51	0.02 \pm 0.46	0.03 \pm 0.90	0.01 \pm 0.58	0.001 \pm 0.39			
	Undetected	1.13	0.95	0.80	0.94	0.89	0.91	0.89	0.91	-0.60 \pm 1.44	-0.08 \pm 1.47	0.31 \pm 1.25	-0.27 \pm 1.49	-0.04 \pm 1.41	0.18 \pm 1.43			
	Overall	0.93 ^b	0.66	0.34	0.79 ^b	0.62	0.34	0.62	0.34	-0.40 \pm 1.56	-0.04 \pm 1.20	0.05 \pm 0.65	-0.16 \pm 1.46	-0.01 \pm 1.14	0.02 \pm 0.65			
Provider 3	Detected	1.19	1.37	1.73	1.12	0.86	1.40	0.86	1.40	0.10 \pm 2.33	0.07 \pm 2.61	0.11 \pm 3.31	0.09 \pm 2.20	0.04 \pm 1.68	0.09 \pm 2.71			
	Undetected	2.02	2.27	8.21	2.00	1.97	6.75	2.00	1.97	0.70 \pm 3.51	0.74 \pm 3.97	3.86 \pm 13.39	0.68 \pm 3.48	0.64 \pm 3.41	2.61 \pm 11.28			
	Overall	1.73	1.75	2.38 ^c	1.69	1.37	2.01	1.69	1.37	0.45 \pm 3.16	0.31 \pm 3.25	0.28 \pm 4.54	0.43 \pm 3.10	0.26 \pm 2.58	0.22 \pm 3.85			

RMSE; root mean square error; LOA; limits of agreement; PGM; Programme feed.

^aSignificant difference ($p < 0.05$) to cam 1 feed for the same provider and resolution.

^bSignificant difference ($p < 0.05$) to tactical camera feed for the same provider and resolution.

^cSignificant difference ($p < 0.05$) between camera resolution for the same provider and camera feed.

Provider 1, 99.7% for Provider 2, and 91.2% for Provider 3 (Figure 3F).

Provider 1 had the most consistent accuracy for speed compared to TRACAB (RMSE = 0.4 to $0.6 \text{ m}\cdot\text{s}^{-1}$). The accuracy of Providers 2 and 3 was more variable across configurations, with some significant differences observed (Table 3). Specifically, for Provider 2, there were larger errors for the programme feed compared to the tactical feed ($p < 0.001$). For Provider 3 tactical feed, the 720p resolution was significantly poorer than the 1080p ($p < 0.001$).

Providers 1 and 2 had the best accuracy when they detected the player from the footage (RMSE: Provider 1 = $< 0.35 \text{ m}\cdot\text{s}^{-1}$; Provider 2 = $< 0.46 \text{ m}\cdot\text{s}^{-1}$). Provider 3's accuracy was also improved for detected data, although the errors were still larger than the other providers. Providers 1 and 2 had the best accuracy for undetected data (RMSE = 0.78 to $1.13 \text{ m}\cdot\text{s}^{-1}$).

Total distance accuracy

The average total distance reported by TRACAB for each player was $7997 \pm 3297 \text{ m}$.

For total distance accuracy (Table 4), there were significant main effects of provider ($p < 0.001$) and camera feed ($p < 0.001$). The best accuracy for total distance offered by Provider 2 was derived from the tactical 1080p feed (mean bias [LOA] = 96 m [-94 to 232 m]). The best accuracy for Provider 1 was using the tactical 1080p feed (mean bias [LOA] = -271 m [-452 to -90 m]), while for Provider 3, it was using the camera 1 1080p feed (mean bias [LOA] = 1163 m [321 to 2005 m]).

Accuracy varied for Providers 1 and 2 across camera feeds, with the programme feed significantly poorer than the camera 1 and tactical feeds ($p < 0.001$). Similarly, for Provider 3, accuracy was significantly reduced for the 1080p programme compared to the camera 1 and tactical feeds ($p < 0.001$).

Mean bias ranged from -297 to 471 m when players were detected regardless of camera feed. When players were undetected, accuracy was compromised for Provider 3 across all camera feeds (mean bias = 804 to 1794 m). This was consistent for Providers 2 and 3 using the programme feed (mean bias = -1745 to -750 m). However, using the tactical and camera 1 feeds resulted in improved accuracy (mean bias = -236 to 84 m).

Discussion

The aim of this study was to: (1) quantify the accuracy of commercially available computer-vision and artificial intelligence (AI) player tracking software to measure player position, speed and distance covered using broadcast footage and (2) determine the impact of camera feed and video resolution on accuracy. It should be made clear that the study's

Table 4. Concurrent validity of player software tracking systems to measure total distance (m) during football match-play in comparison to an optical tracking system.

		Mean bias (m) \pm 95% LOA					
		720p			1080p		
		PGM	Cam 1	Tactical	PGM	Cam 1	Tactical
Provider 1	Detected	-184 \pm 62	-222 \pm 93	-298 \pm 140	-175 \pm 63	-212 \pm 82	-287 \pm 109
	Undetected	-1489 \pm 381	-236 \pm 353	-3 \pm 109	-1490 \pm 384	-231 \pm 350	16 \pm 108
	Overall	-1672 \pm 398 ^{a,b}	-458 \pm 386	-302 \pm 220	-1665 \pm 404 ^{a,b}	-442 \pm 380	-271 \pm 181
Provider 2	Detected	-56 \pm 64	-9 \pm 30	66 \pm 114	44 \pm 102	15 \pm 37	12 \pm 51
	Undetected	-1689 \pm 1107	-150 \pm 456	158 \pm 136	-750 \pm 401	-74 \pm 351	84 \pm 152
	Overall	-1745 \pm 1099 ^{a,b,c}	-159 \pm 469	225 \pm 191	-706 \pm 426 ^{a,b}	-59 \pm 365	96 \pm 163
Provider 3	Detected	150 \pm 267	174 \pm 488	591 \pm 2176	130 \pm 206	109 \pm 230	471 \pm 1190
	Undetected	1794 \pm 1106	1222 \pm 1427	804 \pm 1984	1747 \pm 999	1054 \pm 791	713 \pm 1955
	Overall	1945 \pm 1303	1395 \pm 1775	1395 \pm 4116	1877 \pm 1124 ^{a,b}	1163 \pm 842	1184 \pm 3776

LOA; limits of agreement, PGM; Programme feed.

^aSignificant difference ($p < 0.05$) to cam 1 feed for the same provider and resolution.

^bSignificant difference ($p < 0.05$) to tactical camera feed for the same provider and resolution.

^cSignificant difference ($p < 0.05$) between camera resolution for the same provider and camera feed.

aim was not to appraise the validity of each specific provider, but rather to develop an understanding of how viable computer-vision and AI is for tracking team sport players during competition. The main findings of this study show that the accuracy of computer-vision and AI tracking software for measuring player position and speed is dependent on several factors, including the processing techniques used by the provider, the camera feed, and the video resolution. Moreover, the detection of the player is crucial for accurate tracking to occur. It is likely that this technology will continue to improve as AI techniques and computational power should develop. This study builds on previous research (Mazzeo et al., 2008; Naik and Hashmi, 2021; Stein et al., 2017), which has demonstrated that players can be accurately detected using broadcast-acquired footage but, to date, has not evaluated the accuracy of tracking their position and speed. Overall, this study shows promise in the use of computer-vision and AI player tracking. Further research is warranted to develop data processing, computer-vision and AI standards for which providers can adhere to and maximise the quality of the data, as well as develop a better understanding of what this technology can be used for.

In the context of spatial tracking, position accuracy is not currently suitable (RMSE = 1.68 to 16.39 m), with substantial proportions of positional error above 1.0 m for the programme and camera 1 feeds (Figure 3A and C). Specifically, the proportion of position error under 1.0 m ranged from 24.1 to 69.3% for the programme feed and 60.5 to 63.7% for the camera 1 feed across providers. In contrast, the tactical feed demonstrated improved accuracy, with 82.6 to 93.7% of position error under 1.0 m across providers (Figure 3E). Further, it was shown that, across multiple configurations, accuracy improves significantly when the player is detected (RMSE = 0.44 to 1.14 m).

Therefore, in isolated moments when players are detected by the software (e.g., set play or goal), practitioners could use this data for tactical analysis (e.g., spatial tracking) (Andrienko et al., 2019). In line with this, the relatively higher accuracy observed in the tactical feed (approaching ~95% of positional error < 1.0 m) suggests that, in specific contexts, the data may still be suitable for lower-grade or exploratory applications such as player scouting in competitions where comprehensive tracking data is not routinely available, and where having some imperfect spatial information is often more valuable than having no data at all. As for speed, Providers 1 and 2 showed approximately 87.5 to 99.7% of error less than $1 \text{ m}\cdot\text{s}^{-1}$ (Figure 3B, D and F). Further, a RMSE of $0.34 \text{ m}\cdot\text{s}^{-1}$ is only 3.4% of the peak speed ($10 \text{ m}\cdot\text{s}^{-1}$) likely to be observed in team sport, which may be acceptable relative accuracy for applied use. Overall, despite a high proportion of small errors and a relatively low RMSE, speed estimates remain less precise than GNSS-derived measures (TEE = 0.10 to 0.23) (Crang et al., 2024).

Across providers, there was no clear configuration that consistently produced the best accuracy when measuring position or speed, highlighting that the provider has a significant influence on validity. This is unsurprising given the numerous steps involved in processing the data to track the players (e.g., calibration, filter type, machine learning approaches used) will differ between providers. Prior to assessing player movements, the playing area must first be calibrated to understand the position and dimensions of the pitch relative to the camera (Breytenbach and Grobler, 2025). This may involve various homography techniques which are commonly used in the context of computer-vision (Pandya et al., 2023). Second, the software must then be able to detect individual players, which will typically rely on defining

attributes such as boot colour and playing shirt number. Third, given the number of machine learning algorithms available (e.g., You Only Look Once, Convolutional Neural Networks), provider-developed AI models will differ as well as the way in which model hyperparameters are tuned (Andrews et al., 2024). Fourth, several computer-vision and AI techniques allow for multi-object tracking during match-play, whilst players are in and out of the camera field of view (Cui et al., 2023). Like GNSS providers, this intellectual property is not disclosed to the end-user, and it is therefore difficult to determine the best-practice AI and data processing methods to implement in this ever-evolving field. Given Provider 1 recorded the best validity when using the programme feed to measure position, it would appear they implement AI techniques superior at accounting for changes in camera view and angles. The major difference between providers using this feed appears to come from when players were undetected, with Provider 1 reporting a RMSE of 4.6 m for position during undetected frames, as opposed to 10.0 to 11.6 m for the other providers. This suggests that Provider 1 implemented a superior AI interpolation model that can better estimate the position of a player when outside of the camera's field of view. This appears to be complex however, and may depend on the camera feed used. Overall, position and speed measures do not appear usable in many contexts when the AI model is required to impute player location (e.g., player outside the field of view of the camera). Future research should focus on improving the interpolation models used by the providers to enhance overall accuracy.

Player detection is key in maximising accuracy. This is supported by Providers 1 and 2 who have superior accuracy for position (RMSE = 0.44 to 2.23 m) and speed (RMSE = 0.20 to 0.46 m·s⁻¹) from frames when a player was detected compared to when they were not. Thus, most of the overall error can be attributed to when the player is undetected, further highlighting that refining the interpolation models used to estimate position and speed may enhance overall accuracy. Improving player detection appears just as important, by minimising frames where players are out of the cameras field of view. For example, there is a significant improvement in position validity for Provider 2 when using the tactical feed compared to the programme and camera 1 feeds. The tactical feed provides an elevated and wide angle of the field, increasing the field of view (i.e., increasing player visibility) and limiting occlusion (Harville, 2004). Therefore, the computer-vision tracking software can detect the player for a greater number of frames (89 to 96% detection) compared to the other feeds (36 to 64% detected), again relying much less on the interpolation to estimate position when the player is not detected. Ensuring practitioners can determine whether a player was detected may be useful for tactically analysing isolated periods of play (e.g., opposition set pieces) to gain greater

confidence in the data. If it is found that the player was not detected, this data should not be used in such contexts. Despite showing a significant effect of video resolution on validity, there does not appear to be any practically meaningful changes between video resolution configurations.

While this research focuses on the ability of the software to estimate position and speed, it is important to consider the influence it has on commonly reported metrics such as total distance covered that is derived from these metrics. It appears that Providers 1 and 2 have the best accuracy for measures of total distance when using the camera 1 (percentage error = -0.74 to -5.73%) and tactical feeds (percentage error = -1.20 to 3.78%), while Provider 3 reported poorer accuracy (percentage error = 14.54 to 24.32%). Once again, distance accuracy appears linked to player detection, with player detection greater for the camera 1 and tactical feeds compared to the programme feed. This was highlighted by Provider 2, as most of the error came from situations where a player was undetected. It is important to note that at times while the accuracy of position and speed appeared better for Provider 1 using the programme feed compared to other providers, when the data were aggregated to calculate total distance, there was a large difference compared to TRACAB. This potentially could be explained by the direction of the error, with even small errors in same direction capable of accumulating and magnifying the discrepancy in total distance. In contrast, other configurations may present less accuracy at individual data points, but if this error varies in direction, then the influence on total distance is decreased.

Overall, it is recommended that providers use a tactical feed with 720p or 1080p video resolution when tracking players' position and speed. This method will maximise the number of video frames the players are visible (i.e., in the camera's field of view) and detected for, improving accuracy. This however is reliant upon the provider implementing the correct computer-vision and AI model.

A limitation of this study was that the data was collected from a single match, stadium and broadcaster in what could be considered prime conditions. In turn, stadium design, camera angles and jersey colour may vary at other stadiums which could influence the validity of the software. For example, a stadium with a tactical camera that is at a lower elevation (e.g., narrower field of view) may introduce more undetected data points. Similarly, camera operators or broadcasters may vary in the way that they film the match (e.g., holding a wide-shot vs zooming in and out), which could also result in more undetected data points. Although the study was limited to a single match, simulation and empirical work (i.e., fixed-effect only models) suggest that fixed-effect estimates from linear mixed models are relatively robust under such conditions, whereas uncertainty in variance components may be greater. Importantly, our main conclusions are driven by fixed-effect estimates that were stable across a range of


plausible model structures, supporting the robustness of the findings. Another limitation is that goalkeepers were excluded from the analysis, and therefore the findings cannot be generalised to them until further research is conducted. While beyond the scope of this study, future research should also examine the accuracy of performance indicators provided by these providers (e.g., high-speed running distance and maximal speed).


Conclusions

Players can be tracked with computer-vision and AI software that uses video, though accuracy is heavily dependent on the providers software having a suitable computer-vision and AI model, as well as selecting the correct camera feed and video resolution. It is important to maximise the number of frames in which a player is detected, which is achieved by increasing their visibility by using a tactical feed (i.e., wide-view elevated camera). This is highlighted in this study with player detection much greater using the tactical feed (89 to 96%) compared to the programme and camera 1 feeds (36 to 64%). The AI techniques implemented by the provider appears to have an influence, with some providers showing superior validity compared to others when a player is not detected by the software. Overall accuracy is substantially reduced due to the large errors observed when players are not detected by the software, with these frames showing markedly poorer accuracy than frames in which players are successfully detected. Regardless, future research should focus on refining the interpolation methods implemented by the providers to improve overall accuracy. While there was a significant effect of video resolution on validity, there was no practically meaningful difference between configurations. It is important to also consider the influence of the software's position and speed accuracy on derived variables such as aggregated total distance. For best accuracy, it is recommended providers use a tactical feed with a 720p or 1080p video resolution to track players. Consumers should be aware that validity may change between providers, given they may implement different computer-vision and AI models.

ORCID iDs

Zachary L Crang  <https://orcid.org/0009-0001-4915-4022>

Katie L Mills  <https://orcid.org/0000-0002-1828-7041>

Michael H Cole  <https://orcid.org/0000-0003-1817-7528>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

The data used for this study were collected by FIFA and collaborating broadcast tracking providers. Due to media and data rights, the datasets are not publicly available but can be requested by contacting the authors KM or JB of this article.

Supplemental material

Supplemental material for this article is available online.

References

- Andrews P, Borch N and Fjeld M (2024) FootyVision: Multi-object tracking, localisation, and augmentation of players and ball in football video. In: *Proceedings of the 2024 9th international conference on multimedia and image processing*. Association for Computing Machinery, pp.15–25.
- Andrienko G, Andrienko N, Anzer G, et al. (2019) Constructing spaces and times for tactical analysis in football. *IEEE Transactions on Visualization and Computer Graphics* 27(4): 2280–2297.
- Barris S and Button C (2008) A review of vision-based motion analysis in sport. *Sports Medicine* 38: 1025–1043.
- Breytenbach G and Grobler J (2025) Evaluating the accuracy of a generic field template for camera calibration in soccer broadcast footage. *SN Computer Science* 6(2): 07.
- Clemente FM, Couceiro MS, Martins FML, et al. (2013) Activity profiles of soccer players during the 2010 world cup. *Journal of Human Kinetics* 38: 201–211.
- Crang ZL, Duthie G, Cole MH, et al. (2024) The validity of raw custom-processed global navigation satellite systems data during straight-line sprinting across multiple days. *Journal of Science and Medicine in Sport* 27(3): 204–210.
- Cui Y, Zeng C, Zhao X, et al. (2023) Sportsmot: A large multi-object tracking dataset in multiple sports scenes. In: *Proceedings of the IEEE/CVF international conference on computer vision*. IEEE, pp.9921–9931.
- Delves RI, Duthie GM, Ball KA, et al. (2022) Applying common filtering processes to Global Navigation Satellite System-derived acceleration during team sport locomotion. *Journal of Sports Sciences* 40(10): 1116–1126.
- Duthie G, Pyne D and Hooper S (2003) The reliability of video based time motion analysis. *Journal of Human Movement Studies* 44(3): 259–272.
- Figuroa PJ, Leite NJ and Barros RM (2006) Tracking soccer players aiming their kinematical motion analysis. *Computer Vision and Image Understanding* 101(2): 122–135.
- Gabriel PF, Verly JG, Piater JH, et al. (2003) The state of the art in multiple object tracking under occlusion in video sequences. In: *Advanced Concepts for Intelligent Vision Systems*. Citeseer, pp. 166–173.

- Goes F, Kempe M, Van Norel J, et al. (2021) Modelling team performance in soccer using tactical features derived from position tracking data. *IMA Journal of Management Mathematics* 32(4): 519–533.
- Harville M (2004) Stereo person tracking with adaptive plan-view templates of height and occupancy statistics. *Image and Vision Computing* 22(2): 127–142.
- Hurault S, Ballester C and Haro G (2020) Self-supervised small soccer player detection and tracking. In: *Proceedings of the 3rd international workshop on multimedia content analysis in sports*. Association for Computing Machinery, pp.9–18.
- Iwase S and Saito H (2003) Tracking soccer players based on homography among multiple views. In: *Visual Communications and Image Processing 2003*. SPIE, pp.283–292.
- Iwase S and Saito H (2004) Parallel tracking of all soccer players by integrating detected positions in multiple view images. In: *Proceedings of the 17th international conference on pattern recognition, 2004*. ICPR 2004. IEEE, pp.751–754.
- Jaques T and Pavia G (1974) An analysis of the movement patterns of players in an Australian Rules league football match. *Australian Journal of Science and medicine* 5(10): 10–21.
- Johnston RD, Devlin P, Wade JA, et al. (2019) There is little difference in the peak movement demands of professional and semi-professional rugby league competition. *Frontiers in Physiology* 10: 1285.
- Johnston RD, Thornton HR, Wade JA, et al. (2020) The distribution of match activities relative to the maximal mean intensities in professional rugby league and Australian football. *The Journal of Strength and Conditioning Research* 10: 1360–1366.
- Linke D, Link D and Lames M (2020) Football-specific validity of TRACAB's optical video tracking systems. *PLoS One* 15(3): e0230179.
- Mazzeo PL, Spagnolo P, Leo M, et al. (2008) Visual players detection and tracking in soccer matches. In: *2008 IEEE fifth international conference on advanced video and signal based surveillance*. IEEE, pp.326–333.
- Modric T, Versic S and Sekulic D (2020) Position specific running performances in professional football (soccer): Influence of different tactical formations. *Sports* 8(12): 61.
- Naik BT and Hashmi MF (2021) Ball and player detection & tracking in soccer videos using improved YOLOv3 model.
- Nettleton B and Sandstrom E (1963) Skill and conditioning in Australian rules football. *The Australian Journal of Physical Education* 29: 17–30.
- Omidshafiei S, Hennes D, Garnelo M, et al. (2022) Multiagent off-screen behavior prediction in football. *Scientific Reports* 12(1): 8638.
- Pandya Y, Nandy K and Agarwal S (2023) Homography based player identification in live sports. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, pp. 5209–5218.
- Parker RA, Weir CJ, Rubio N, et al. (2016) Application of mixed effects limits of agreement in the presence of multiple sources of variability: Exemplar from the comparison of several devices to measure respiratory rate in COPD patients. *PLoS One* 11(12): e0168321.
- Schielzeth H, Dingemanse NJ, Nakagawa S, et al. (2020) Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution* 11(9): 1141–1152.
- Scott A, Uchida I, Onishi M, et al. (2022) SoccerTrack: A dataset and tracking algorithm for soccer with fish-eye and drone videos. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, pp.3569–3579.
- Stein M, Janetzko H, Lamprecht A, et al. (2017) Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE transactions on Visualization and Computer Graphics* 24(1): 13–22.
- Thinh NH, Son HH, Dzong CTP, et al. (2019) A video-based tracking system for football player analysis using Efficient Convolution Operators. In: *2019 international conference on advanced technologies for communications (ATC)*. IEEE, pp.149–154.
- Thoseby B, Govus AD, Clarke AC, et al. (2023) Peak match acceleration demands differentiate between elite youth and professional football players. *PLoS One* 18(3): e0277901.
- West SW, Clubb J, Torres-Ronda L, et al. (2021) More than a metric: How training load is used in elite sport for athlete management. *International Journal of Sports Medicine* 42(04): 300–306.
- Whitehead S, Till K, Weaving D, et al. (2019) Whole, half and peak running demands during club and international youth rugby league match-play. *Science and Medicine in Football* 3(1): 63–69.
- Xu M, Orwell J and Jones G (2004) Tracking football players with multiple cameras. In: *2004 International Conference on Image Processing*. ICIP'04. IEEE, pp.2909–2912.