Semi-automated last touch detection for out-of-bounds possession decisions in football

WANG, Henry, MILLS, Katie, BILLINGHAM, Johsan, ROBERTSON, Sam and HOSOI, A. E.

Citation:

WANG, Henry, MILLS, Katie, BILLINGHAM, Johsan, ROBERTSON, Sam and HOSOI, A. E. (2025). Semi-automated last touch detection for out-of-bounds possession decisions in football. Sports Engineering, 28 (2), p. 36. [Article]

Sports Engineering (2025) 28:36
https://doi.org/10.1007/s12283-025-00518-3

METHODOLOGY

Semi-automated last touch detection for out-of-bounds possession decisions in football

Henry Wang[1] · Katie Mills[2,3] · Johsan Billingham[2] · Sam Robertson[4,5] · A. E. Hosoi[1]

Received: 7 April 2025 / Revised: 8 August 2025 / Accepted: 14 August 2025

Abstract

Football referees must make quick and accurate decisions in unforgiving environments. In parallel, advances in optical tracking have created new avenues for technology-assisted officiating. Using skeletal and ball tracking data, we present a novel diphase framework for Semi-automated Last Touch detection, designed to help referees adjudicate out-of-bounds possession decisions where player and ball occlusion may pose challenges. The proposed methodology uses a touch probability model to find the decision frame of the last touch before the ball goes out-of-bounds, and rules-based or supervised learning algorithms predict the player responsible for the touch. Leveraging principles of kinematics, human anthropometry, and machine learning, the models predict the correct possession decision with up to 82.5% accuracy on a test dataset of duels from the 2022 FIFA World Cup, including over 90% for aerial duels. Our results represent potential improvements in human performance reported in previous literature and provide a baseline benchmark for future studies.

Keyword Technological aids, Referees, Decision-making

1 Introduction

The primary role of a football referee is to adjudicate decisions based on *The Laws of the Game* (LOTG) as defined by *The International Football Association Board* (IFAB). Officials are expected to make calls quickly and accurately, often in unfavorable conditions. Complications include rapid speed of play, obstructed visual perception, and noisy environments with potentially contentious managers and players [1, 2]. Studies on professional referees using independent expert refereeing panels report accuracy rates of 0.50 to 0.93 for booking and foul decisions, highlighting the challenging decision-making environments faced [3–6]. The appeal of technology-assisted officiating lies in its perceived objectivity and visual robustness, as tracking data can quantitatively explain events in ways humans cannot. In addition, technology, such as the Video Assistant Referee (VAR), can help increase the accuracy of decisions [7]. However, extended officiating delays associated with some existing tech-assisted systems have drawn criticism from fans and media, emphasizing the need for low-latency solutions [8–10].

Technology's value in the analysis of football matches continues to grow. Collection of large volumes of game footage and annotations, such as the SoccerNet datasets, has facilitated player and ball detection and action spotting, greatly expanding the ability to quantify on-pitch events using computer vision [11–13]. Additionally, previous work has been done in using video of incidents to train models that predict officiating decisions like fouls and cards, even accompanied by explanations [14, 15].

Beyond single-source video, advancements in multi-camera optical tracking now allow for near-real-time delivery of ball and skeletal tracking data, produced by an ensemble of in-stadium cameras. Balls can be embedded with an inertial measurement unit (IMU) and ultra-wideband sensors, allowing the implementation of these technologies to aid officials. In 2012, the Fédération Internationale de Football Association

✉ Henry Wang
hwang21@mit.edu

[1] MIT Sports Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

[2] Fédération Internationale de Football Association, Zürich, Switzerland

[3] Sports Engineering Research Group, Sheffield Hallam University, Sheffield, UK

[4] Institute for Health and Sport (IHeS), Victoria University, Melbourne, VIC, Australia

[5] School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, QLD, Australia

Published online: 27 August 2025

(FIFA) introduced goal line technology to detect when the ball crosses the goal line automatically. Recent FIFA World Cups introduced Semi-automated Offside Technology (SAOT), leveraging skeletal and ball tracking data to transition offside decisions into a faster, more efficient system.

Last-touch detection for out-of-bounds (OOB) scenarios remains an unexplored area of tech-assisted refereeing in football but merits attention for two key reasons. First, OOB scenarios result in a possession decision, which has been shown to impact team behavior and performance [16–18]. Second, and perhaps more importantly, OOB scenarios at the goal lines determine whether play is resumed with a corner kick or a goal kick. The offensive value of corner kicks is well-documented in the literature. Analyses of elite international tournaments indicate that corners led to a team leveling the score or taking the lead in 73% of cases, and contributed to winning or drawing the match 76% of the time [19, 20]. Corner kicks in the 2017–2018 FA Women's Super League season resulted in 38 goals, accounting for 13.5% of the total 282 goals scored [21]. In these match-altering moments, exploring technological tools to assist the referee is well motivated.

This paper proposes a novel approach for Semi-automated Last Touch detection (SALT) using optical tracking data suitable for low-latency operation. The proposed framework operates in two phases. First, ball tracking data is run through a touch probability model that produces a decision frame, or the instant when the last touch is detected. Then, four metrics are engineered at the decision frame. These metrics can be used in rules-based approaches or aggregated into a binary classifier that predicts which player was responsible for the touch. In this paper, the rationale behind each step in the methodology is detailed, and SALT performance is evaluated on a dataset of challenging duel scenarios with ground truth labels of who touched the ball last.

We summarize our contributions as follows. First, we propose a diphase framework for SALT that operates on skeletal and ball tracking data. Second, we present multiple approaches for inferring the responsible player under this framework, combining rules-based and machine learning methods. Finally, we evaluate these methodologies on a dataset of challenging duels, provide a formal recommendation with supporting rationale, and demonstrate the feasibility of last-touch detection using tracking data, establishing both a baseline and directions for future research.

2 Data and methods

2.1 Data resources

A collection of 330 videos from broadcast footage (Duration: 5 s, Resolution: 1080p) capturing 2-opposition player duels from FIFA World Cup 2022 (FWC22) matches was manually annotated by a FIFA researcher (KM). The inter-annotator reliability was assessed across 30 randomly selected videos (HW). This dataset included a range of scenarios in which the ball either remained in play or went OOB following the duel. For each clip, the researcher was given the identities of the two players involved and was tasked with identifying the player and selecting the body part that made the final contact with the ball. In instances where it was difficult to distinguish the player or body part, the duel was marked as non-visible. A majority (195 of 330) were aerial duels involving the head, 109 occurred with the legs or feet, and 26 occurred with the torso, arms, or an unclear body part.

Skeletal and ball tracking data for all 330 duels were provided by Hawk-Eye Innovations (HEI) Player and Ball Tracking System [22] using high-definition multi-camera optical tracking. This technology was tested and certified by the FIFA Quality Programme in September 2021, with details in the Online Resource. The system tracks the xyz position of 29 joints on each player's body and the ball at 50 Hz, which we use to detect contact with the ball. The skeletal data is described in greater detail in section 2.2.

We identify two key challenges with the use of skeletal data for SALT. First, tracking quality and precision may decrease in close opposition scenarios, resulting in inaccurate player and ball positions. Second, only a subset of the 29 joints are directly tracked, while extremities (e.g., head, toes, heels, fingers, hips) are extrapolated. This work attempts to develop a framework using both distance-dependent and distance-independent metrics to enhance the robustness to tracking uncertainty.

All matches during FWC22 used the Adidas Connected Official Match Ball equipped with Kinexon Connected Ball Technology [23], which we refer to as an instrumented ball/football. This ball, with an embedded IMU sensor, samples movement data at 500 Hz, allowing for the extraction of touch timestamps to the nearest 0.002 s that were down-sampled and time-synchronized with tracking data. We also propose a touch probability algorithm, detailed in section 2.2.2, that can predict touches in the absence of the instrumented ball, offering a more accessible alternative, which becomes the focus of this paper. The instrumented ball was tested and certified by the FIFA Quality Programme in September 2021, with details in the Online Resource.

For each duel, we had access to a 5-second window segment of skeletal and ball tracking data. However, the exact duel frames were unknown but essential for analysis. To identify the duel frames within this window, heuristics using two thresholds were used.

1. The mid-hips of both players are within 1.25 m apart.
2. At least one player's mid-hip is within 0.75 m of the ball.

The first threshold ensures physical proximity between players, and the second stipulates that the scuffle is occurring on the ball. These thresholds were applied as a filter across the tracking data segment, and the longest consecutive stretch of remaining frames was taken. For this duel dataset, we found that using the last 25 frames captured the last touch. In real-world implementation, a manual operator may verify that the tracking frames used contain the last touch.

2.2 Methods

In this section, we detail the methods behind each piece of the SALT framework, from how players are represented to how a final last touch prediction is made.

2.2.1 Player representations

The player representations are first defined, as summarized in Fig. 1.

The skeleton was augmented by extrapolating the coordinates of each player's crown, a key part of aerial duels (shown in Fig. 1(b)). Let $\vec{r}_j \in \mathbb{R}^3$ be the position of joint j. First, the midpoint of the eyes ($\vec{r}_{\text{mid-eyes}}$) was computed using the positions of the left and right eyes.

Next, the unit vector pointing from the nose to the midpoint of the eyes was defined:

$$\vec{u} = \frac{\vec{r}_{\text{mid-eyes}} - \vec{r}_{\text{nose}}}{\left\| \vec{r}_{\text{mid-eyes}} - \vec{r}_{\text{nose}} \right\|} \tag{1}$$

Finally, the crown coordinate was computed by extending the vector from the nose in the direction of $\vec{u}$ by 0.18 m based on mean measurements of the human head [24]:

$$\vec{r}_{\text{crown}} = \vec{r}_{\text{nose}} + 0.18 \cdot \vec{u} \tag{2}$$

In addition to estimating $\vec{r}_{\text{crown}}$, the locations of the left and right thumbs and fifth digit were pruned because the duel dataset did not include any handballs.

The skeletal representation models each athlete as a collection of 30 limb connections, or line segments connecting joints on the body. Not all connections, specifically in the

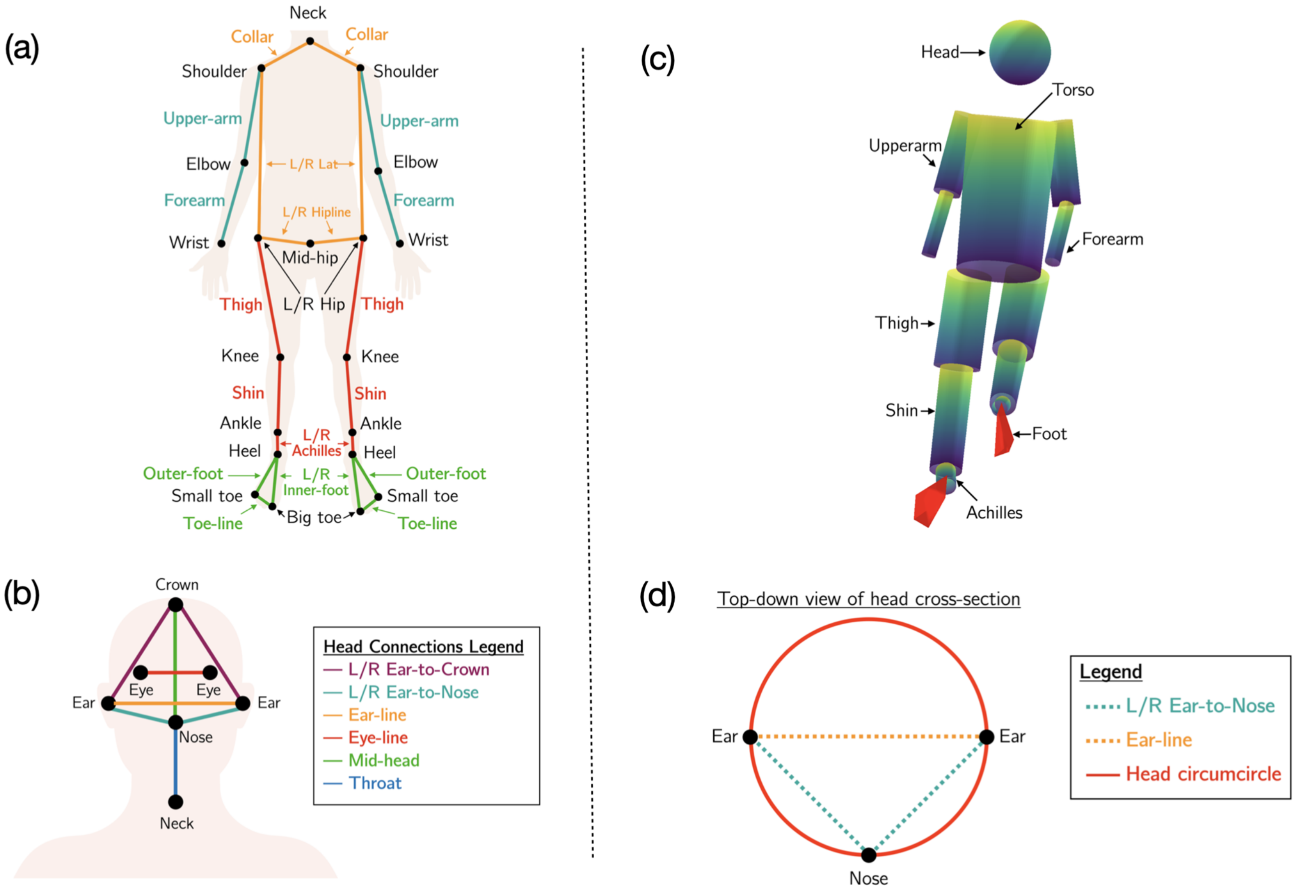


Fig. 1 **(a)** Skeletal coronal representation of tracked landmarks comprising all joints and limb connections, excluding the head **(b)** Skeletal representation of tracked landmarks comprising all joints and limb connections in the head **(c)** Volumetric representation of all body parts **(d)** Transverse-plane view illustrating how the head sphere of the volumetric representation is formed

head, are limbs in the anatomic sense. However, limbs capture the space between joints where the body may contact the ball. Figure 1(a) reports the set of limbs used in the skeletal representation and the joints used to connect each limb. Several quasilimbs are drawn in the head (scalp, eyeline, earline, face, midhead) to improve header detection, shown in Fig. 1(b).

A limitation of the skeletal representation is its lack of volume, which the volumetric representation seeks to address using a collection of spheres, triangular prisms, and cylinders to create a more realistic player model, depicted in Fig. 1(c). The head is approximated by a sphere formed from the circumcircle of the triangle that connects the left ear, the right ear, and the nose, with the circumcircle serving as the bisecting cross section (Fig. 1 (d)). The feet are modeled as triangular prisms, with a base formed by the heel, the big toe, and the small toe, and a height of 0.063 m based on measurements of a popular football foot (US size 11). The limbs are represented by right circular cylinders with radii derived from mean anthropometric data of elite Spanish footballers [25]. The forearms (radius 0.027 m) connect the elbow and wrist, the upper arms (radius 0.052 m) connect the shoulder and elbow, the thighs (radius 0.085 m) connect the hip and knee, and the shins (radius 0.06 m) connect the knee and ankle. The Achilles tendon is also modeled as a cylinder (radius 0.037 m) that connects the ankle and heel. Finally, the ball is represented as a sphere with a radius of 0.11 m, consistent with Law 2 of the *LOTG* [26].

2.2.2 The decision frame

To assert which player touched the ball last, we identify the predicted decision frame, $\hat{f}_d$, which is the last instant when the ball was touched prior to going OOB. To find $\hat{f}_d$, we estimate the sequence $\{p^f_{\text{touch}}\}^{25}_{f=1}$, which represents the probability of ball contact over the 25 consecutive frames indexed by f. The *last* peak in this sequence is $\hat{f}_d$, as there can be multiple peaks in the 25 frame sequence.

We engineer a feature space to maximize linear separability and train a binary classifier to predict p^f_{touch} at any frame f using HEI ball tracking as inputs and instrumented football touches as labels. The first feature is Velocity Cosine Similarity (VCS), defined as the cosine similarity between the current velocity vector at frame f and the outgoing velocity vector at $f + 1$. For the boundary frames $f = 1$ and $f = 25$, p^f_{touch} is estimated using tracking data from frames $f = 0$ and $f = 26$, respectively, which are defined and contain accessible data but excluded from the 25-frame duel sequence. Intuitively, a VCS of 1 indicates a low probability of ball touch, as the trajectory remains undisturbed. If VCS is −1, the probability of ball touch is high, as the ball has begun traveling back toward its origin.

Definition 1 Velocity Cosine Similarity (VCS): Let $\vec{v}_f$ be the instantaneous velocity vector at frame f and $\vec{v}_{f+1}$ be the outgoing velocity vector at frame $f + 1$.

$$VCS := \frac{\vec{v}_f \cdot \vec{v}_{f+1}}{\|\vec{v}_f\| \|\vec{v}_{f+1}\|} \in [-1, 1]$$

The second feature is Velocity Magnitude Ratio (VMR), which captures changes in speed of the ball and ranges between 0 and 1. VMR approaches zero as the current and outgoing velocity vectors become increasingly different in magnitude. When VMR is close to 1, the velocity of the ball exhibits minimal change between frames, indicating a low touch probability.

Definition 2 Velocity Magnitude Ratio (VMR): Let $\vec{v}_f$ be the instantaneous velocity vector at frame f and $\vec{v}_{f+1}$ be the outgoing velocity vector at frame $f + 1$.

$$VMR := \min\left\{ \frac{\|\vec{v}_f\|}{\|\vec{v}_{f+1}\|}, \frac{\|\vec{v}_{f+1}\|}{\|\vec{v}_f\|} \right\} \in [0, 1]$$

Finally, the last feature is minimum closeness, simply the Euclidean distance between the ball and the nearest limb on any player at a given frame. Intuition says frames where minimum closeness is high are unlikely to contain a touch.

Using the three predefined features, we train a logistic regression classifier $\hat{\mathbb{P}}_{touch}$ with L2 regularization using a 70/30 train-test split. There were a total of 7807 unique frames of ball tracking data that were down-sampled and synced with the touch data at 50 Hz. The dataset was highly imbalanced, as about 5% of frames were touches. A balanced class weight loss function was employed in training, which over-penalized false negatives, and 5-fold cross validation optimized the feature polynomial order. The touch model achieves an out-of-sample ROC AUC of 0.97 and F1-score of 0.71, and exhibits low precision as a result of the class-balanced loss function, but very high recall, ensuring no touches are missed.

To find $\hat{f}_d$, we define a touch probability sequence outputted by $\hat{\mathbb{P}}_{touch}$ over the duel frames and found the *last* peak in probability exceeding the height threshold $h_{thresh} = 0.75$ over the last 25 frames, which was selected using a grid search. This process could be considered a temporal feature selection step to determine which time slice should be used to make predictions. Details of the selection of h_{thresh} are found in the Online Resource. We encourage the reader to review section 2.2.3 beforehand for a proper context.

Definition 3 Predicted Decision Frame ($\hat{f}_d$): Let $\hat{p}^f_{touch} = \hat{\mathbb{P}}_{touch}(touch \mid f)$ be the predicted probability of ball touch at frame f. Let $S = \{\hat{p}^f_{\text{touch}}\}^{25}_{f=1}$ be the touch probability sequence. The predicted decision frame is defined as follows.

$$\hat{f}_d = \max\left\{ f \mid f \in \{1,2,\ldots,25\},\ \hat{p}^{f}_{touch} > h_{thresh} \ \wedge\ \hat{p}^{f-1}_{touch} < \hat{p}^{f}_{touch} > \hat{p}^{f+1}_{touch} \right\}$$

If no such frame exists, the decision frame is at the maximum probability in S.

$$\hat{f}_d = \arg\max_{f}\{S\}$$

The decision frame offers flexibility by allowing manual adjustments when deemed incorrect, mirroring the approach used in SAOT and motivating the term "*Semi-automated* last touch."

2.2.3 Rules-based approaches

In this subsection, we propose several rules-based approaches using metrics derived from skeletal and kinematic principles, designed to intuitively infer duel outcomes. The naive closeness approach (NCA) is the "laziest", baseline approach. NCA omits the decision frame and observes the minimum distance between the skeletal representation and the ball over the entire duel, asserting that the player with the smallest minimum distance is the last-touch player. NCA does not distinguish closeness from contact and is blind to signals from relative player and ball movement.

Closeness, the most intuitive approach, measures each pairwise limb-to-ball distance for a given player at $\hat{f}_d$, using the skeletal representation rather than the volumetric representation, and takes the minimum value.

Definition 4 Closeness: For a player p with limbs S_p, let $d_{skel}(\vec{s}_i)$ be the minimum Euclidean distance between a limb $\vec{s}_i \in S_p$ and the ball. The player's closeness, C_p, is:

$$C_p := \min\{d_{skel}(\vec{s}_i) \mid \vec{s}_i \in S_p\}.$$

The closeness approach (CA) asserts that the skeletal representation closest to the ball at $\hat{f}_d$ belongs to the player who touched the ball last. CA represents an improvement over NCA using the decision frame, distinguishing closeness from contact, but does not leverage volumetric characteristics.

Volumetric closeness is analogous to closeness but replaces the skeletal representation with the volumetric representation. We measure the distance between the ball sphere and each boundary surface on the player and then take the minimum value. The volumetric closeness approach (VCA) is identical to the closeness approach but uses volumetric closeness. Additional computational details for $d_{skel}(\cdot)$ and $d_{vol}(\cdot)$ are found in Section 6.

Definition 5 Volumetric Closeness: For a player p with body part shapes B_p, let $d_{vol}(\vec{b}_i)$ be the minimum Euclidean distance between the surface boundary of $\vec{b}_i \in B_p$ and the ball. The player's volumetric closeness, VC_p, is:

$$VC_p := \min\{d_{vol}(\vec{b}_i) \mid \vec{b}_i \in B_p\}.$$

Although closeness is the fundamental indicator of ball contact, it may be unreliable when tracking quality decreases. Under these conditions, we might infer which player kicks the ball by observing the relative motion of the feet and ball. If both are moving in the same direction, then contact is more likely. Under poor tracking, this signal may be more reliable than a distance-dependent metric. Following this intuition, we define joint velocity similarity (JVS) as the cosine similarity between the nearest joint's velocity and the ball's velocity at the decision frame.

Definition 6 Joint Velocity Similarity (JVS): Let $\vec{v}_{ball} \in \mathbb{R}^3$ be the velocity vector of the ball. Let $\vec{v}_{j1}, \vec{v}_{j2} \in \mathbb{R}^3$ be the velocity vectors of the two joints connecting the limb closest to the ball. We define JVS as follows.

$$JVS := \max\left\{ \frac{\vec{v}_{ball} \cdot \vec{v}_{j1}}{\|\vec{v}_{ball}\|\|\vec{v}_{j1}\|}, \frac{\vec{v}_{ball} \cdot \vec{v}_{j2}}{\|\vec{v}_{ball}\|\|\vec{v}_{j2}\|} \right\} \in [-1, 1]$$

Values close to 1 suggest a higher probability of touch. We illustrate this intuition in Fig. 2. Under the Joint Velocity Similarity Approach (JVSA), the player with the higher JVS at the decision frame is assigned the last touch. Physically, a touch is an elastic collision between body and ball and observing a projectile's pre- and post-collision trajectories may help infer the orientation of the collision surface. We leverage this intuition by defining Expected Trajectory Similarity (ETS) using volumetric representations.

At frame $f_d - 1$, we calculate the pre-collision normal vector $\vec{n}_{pre}$ connecting the nearest point on the surface of the closest shape at f_d to the ball for a given player. The calculation of $\vec{n}_{pre}$ depends on whether this shape is a sphere, a cylinder, or a triangular prism. For spheres, the pre-collision normal vector to the ball $\vec{n}^{sphere}_{pre}$ is the vector connecting the origins of both spheres, $\vec{r}_{ball}$ and $\vec{r}_{sphere}$.

$$\vec{n}^{sphere}_{pre} = \vec{r}_{ball} - \vec{r}_{sphere} \qquad (3)$$

For cylinders, we find the normal vector by projecting the ball's position onto the cylinder's axis to get $\vec{r}_{proj}$. The normal vector $\vec{n}^{cyl}_{pre}$ is the vector that connects the origin of the ball to this projection minus the radius of the ball. More

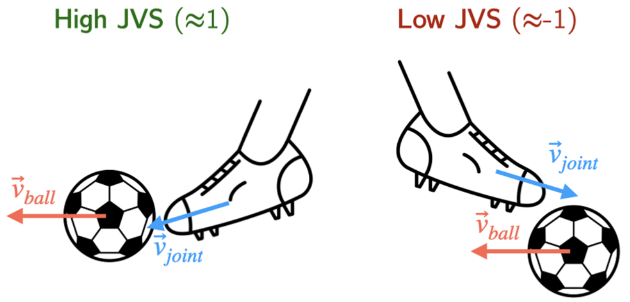


Fig. 2 Illustrating joint velocity similarity (JVS) with joint and ball velocity vectors: On the left, JVS is high as both vectors point in similar directions; on the right, JVS is low as both vectors point in opposite directions

formally, let $\vec{r}_{start}$ and $\vec{r}_{end}$ be the positions of the start and end of the cylinder's axis. Let $\vec{r}_{axis} = \vec{r}_{end} - \vec{r}_{start}$. We compute $\vec{n}^{cyl}_{pre}$ as follows:

$$\vec{n}^{cyl}_{pre} = \vec{r}_{ball} - \left(\vec{r}_{start} + \frac{(\vec{r}_{ball} - \vec{r}_{start}) \cdot \vec{r}_{axis}}{\vec{r}_{axis} \cdot \vec{r}_{axis}} \times \vec{r}_{axis}\right) \tag{4}$$

Finally, for triangular prisms, we use the face of the prism deemed to have touched the ball, either a triangle or rectangle, and we can find the normal vector $\vec{n}^{prism}_{pre}$ by taking the cross product of any two connected edges $\vec{e}_1$ and $\vec{e}_2$, and then orienting the vector so it points towards the ball by ensuring the dot product between $\vec{n}^{prism}_{pre}$ and $\vec{r}_{ball} - \vec{n}^{prism}_{pre}$ is positive. In Eq. 5 below, the "$*$" operator represents scalar multiplication.

$$\vec{n}^{prism}_{pre} = \left(2 * \mathbb{1}\left\{(\vec{e}_1 \times \vec{e}_2) \cdot (\vec{r}_{ball} - \vec{e}_1 \times \vec{e}_2) > 0\right\} - 1\right) * (\vec{e}_1 \times \vec{e}_2) \tag{5}$$

Finally, using $\vec{n}_{pre}$, we define ETS as the cosine similarity between $\vec{n}_{pre}$ and $\vec{v}^{ball}_{post}$, the ball's post-collision velocity at frame $f_d + 1$.

Definition 7 Expected Trajectory Similarity (ETS): Let $\vec{n}^{S}_{pre}$ be the pre-collision normal vector between the nearest-limb shape S and the ball at frame $f_d - 1$. Let $\vec{v}^{ball}_{post}$ be the velocity vector of the ball at frame $f_d + 1$.

$$ETS := \frac{\vec{n}^{S}_{pre} \cdot \vec{v}^{ball}_{post}}{\|\vec{n}^{S}_{pre}\| \|\vec{v}^{ball}_{post}\|} \in [-1, 1]$$

ETS takes values between -1 and 1, with a value of 1 indicating a high collision likelihood. Under the Expected Trajectory Similarity Approach (ETSA), the player with the highest ETS at the decision frame is deemed responsible for the last touch Fig. 3.

We summarize the set of rules-based approaches in Table 1.

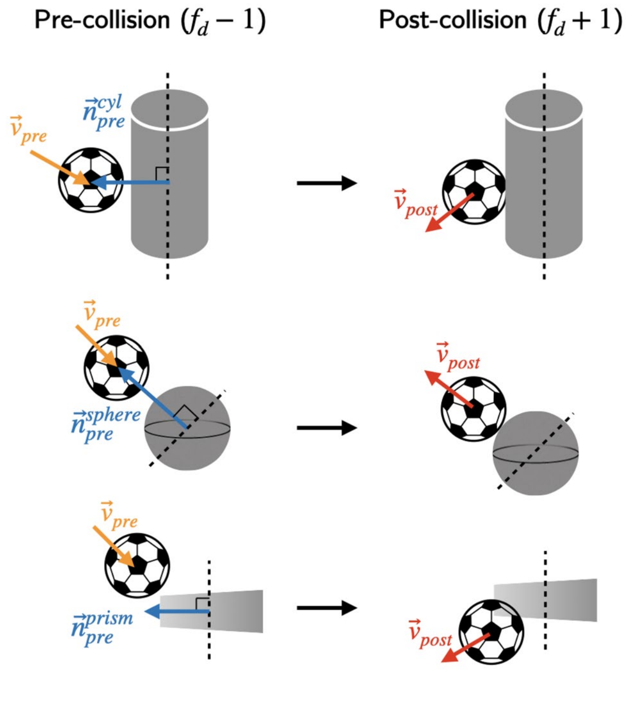


Fig. 3 Illustration of vectors in ETS computation for all possible volumetric shapes. Note that for the prism, a rectangular face is shown, but the same logic applies to triangular faces

2.2.4 Multimodal approaches

The rules-based approaches in section 2.2.3 capture different kinematic signals, potentially offering a balance of predictive power and robustness to tracking uncertainty. We explore developing multimodal predictors that aggregate all four methods. In Table 2, we summarize the five multimodal approaches proposed and detail them in this section, with accuracies reported in section 3. Traditional ensemble methods use unweighted majority or plurality voting to aggregate the outputs of disjoint classifiers, which has been shown to improve overall predictive performance [27, 28]. In our binary classification setting, we implement majority voting on the intermediate decisions from CA, VCA, JVSA, and ETSA, using VCA as the tiebreaker. This Majority Vote Classifier (MVC) serves as the baseline multimodal method.

Definition 8 Majority Vote Classifier (MVC) Define $h_A(\mathbf{x}) \in \{0, 1\}$ as the binary last touch decision produced by rules-based approach A on duel $\mathbf{x}$. The decision $\hat{D}_{MVC}$ is:

$$\hat{D}_{MVC} = \text{sign}\left(\sum_{A} (2h_A(\mathbf{x}) - 1)\right)$$

where each $h_A(\mathbf{x})$ is mapped from $\{0, 1\}$ to $\{-1, +1\}$, and the sum is taken over all classifiers in the set $\{\text{CA, VCA, JVSA, ETSA}\}$.

Table 1 Summary of Rules-based Approaches

Approach	Summary
Naive Closeness Approach (NCA)	The player with the smallest skeletal-to-ball distance over the 25 duel frames is assigned the last touch
Closeness Approach (CA)	The player with the smallest skeletal-to-ball distance at the decision frame is assigned the last touch
Volumetric Closeness Approach (VCA)	The player with the smallest volumetric-to-ball distance at the decision frame is assigned the last touch
Joint Velocity Similarity Approach (JVSA)	The player with the higher Joint Velocity Similarity (JVS) at the decision frame is assigned the last touch
Expected Trajectory Similarity Approach (ETSA)	The player with the higher Expected Trajectory Similarity (ETS) at the decision frame is assigned the last touch

Table 2 Summary of Multimodal Approaches

Approach	Inputs	Model(s)
Majority Vote Classifier (MVC)	Rules-based decisions	Rules-based majority vote
Stacking Classifier (SC)	Rules-based decisions	Binary-input LR
Body-segmented Stacking Classifier (BSSC)	Rules-based decisions	3 x Binary-input LR
Margin Classifier (MC)	Numerical margins	LR
Body-segmented Margin Classifier (BSMC)	Numerical margins	3 x LR

We extend MVC by training ML models to learn the optimal weights for each rules-based output. To reduce variability from train-test splits, we use Monte Carlo cross validation (MCCV) with 500 rounds. A key extension of the multimodal ML approach is the ability to allow the relative importance of inputs to vary depending on the body region involved through segmentation modeling. We determine the duel's body region by identifying, via our closeness metrics, the skeletal and volumetric parts closest to the ball for both players, mapping these parts to one of three regions (upper, lower, or head), and assigning the duel the most frequent region. The Online Resource section contains details of how limbs are mapped to body regions.

We first implement a logistic regression (LR) classifier using binary rule-based decisions as inputs, selecting a linear or quadratic feature space through 5-fold cross-validation. We experimented with more expressive model classes, but found similar performance with loss of interpretability, a key pillar of existing frameworks for evaluating sports technology quality, which includes tools related to tracking systems [29]. This meta-learner, referred to as a Stacking Classifier (SC), employs stack generalization, effective when base classifiers excel in different subspaces [30]. We also implement a Body-segmented Stacking Classifier (BSSC), training separate LR models for duels grouped by body region.

Definition 9 Stacking-based Classifiers: Define $h_A(\mathbf{x}) \in \{0, 1\}$ as the binary last touch decision produced by classifier A on duel $\mathbf{x}$. $\sigma(\cdot)$ is the sigmoid function. Let β be learned weights. For clarity, we only define the first-order specifications here.

1. **Stacking Classifier (SC):** The decision made by the stacking classifier is:

$$\hat{D}_{SC} = \mathbb{1}\left\{\sigma\left(\boldsymbol{\beta}^{\top}\mathbf{h}(\mathbf{x})\right) \geq 0.5\right\}$$

where $\mathbf{h}(\mathbf{x}) = [h_{CA}(\mathbf{x}), h_{VCA}(\mathbf{x}), h_{JVSA}(\mathbf{x}), h_{ETSA}(\mathbf{x})]^{\top}$.

2. **Body-segmented Stacking Classifier (BSSC):** Separate models are trained for each body region. The decision for each region's model is:

$$\hat{D}_{BSSC,\text{region}} = \mathbb{1}\left\{\sigma\left(\boldsymbol{\beta}_{\text{region}}^{\top}\mathbf{h}_{\text{region}}(\mathbf{x})\right) \geq 0.5\right\}$$

where region $\in$ {upper, lower, head}.

One limitation of MVC, SC, and BSSC is the homogeneity of inputs, which do not encode how "close" the decision may be. To address this, we define $\mathbf{m}(\mathbf{x}) = \{m_C, m_{VC}, m_{JVS}, m_{ETS}\}$ as the set of margins produced by CA, VCA, JVSA, and ETSA, respectively, for duel x. More formally, let C_p, VC_p, JVS_p, and ETS_p denote player p's closeness, volumetric closeness, joint velocity similarity, and expected trajectory similarity, respectively. For a duel between players i and j, we compute the components of $\mathbf{m}(\mathbf{x})$ as follows:

$$m_C = C_j - C_i \tag{6}$$

$$m_{VC} = VC_j - VC_i \tag{7}$$

$$m_{JVS} = JVS_i - JVS_j \tag{8}$$

$$m_{ETS} = ETS_i - ETS_j \tag{9}$$

We implement the same two model specifications in the previous section, but replace the binary inputs with $\mathbf{m}(x) \in \mathbb{R}^4$. We call the two methods Margin Classification (MC) and Body-segmented Margin Classification (BSMC), which are analogous to SC and BSSC, respectively.

Definition 10 Margins-based Classifiers: Let $\mathbf{m}(\mathbf{x})$ be the margins vector derived from duel $\mathbf{x}$. $\sigma(\cdot)$ is the sigmoid function. For clarity, we only define the first-order specifications here.

1. **Margin Classifier (MC):** The decision made by the margin classifier is:

   $$\hat{D}_{MC} = \mathbb{1}\left\{\sigma\left(\boldsymbol{\beta}^{\top}\mathbf{m}(\mathbf{x})\right) \geq 0.5\right\}$$

   where $\mathbf{m}(\mathbf{x}) = [m_C, m_{VC}, m_{JVS}, m_{ETS}]^{\top}$.
2. **Body-segmented Margin Classifier (BSMC):** Separate models are trained for each body region. The decision for each region's model is:

   $$\hat{D}_{BSMC,\text{region}} = \mathbb{1}\left\{\sigma\left(\boldsymbol{\beta}_{\text{region}}^{\top}\mathbf{m}_{\text{region}}(\mathbf{x})\right) \geq 0.5\right\}$$

   where region $\in$ {upper, lower, head}.

3 Results

3.1 Decision frame performance

The touch probability model achieved an out-of-sample ROC AUC of 0.97 and an F1-score of 0.71. Notably, the upweighting of positive labels in training resulted in a substantial disparity between precision and recall (0.32 vs. 0.99, respectively). However, a more informative evaluation of model performance considers the temporal difference between the predicted and ground truth decision frames for each duel. This directly captures temporal precision, which is a more meaningful indicator of performance in this context.

Figure 4 depicts the empirical distribution of $\Delta f = f_d - \hat{f}_d$, where f_d is the labeled touch frame from the instrumented ball and $\hat{f}_d$ is the predicted decision frame from the touch probability model. The highest magnitude difference is 13 frames. Manual investigation revealed that large differences were caused by missed instrumented ball touches, errant ball tracking data, and instances where bounces on the pitch were flagged as touches. However, 66.7% of the time $\Delta f = 0$, and $\Delta f \in [-1, 1]$ for 91% of the 301 duels. In real-world implementation, any frame decision errors can be corrected by the operator, which is already standard practice in SAOT when detecting the kick point.

3.2 Rules-based approach results

Figure 5 shows rules-based accuracy differences, with VCA performing best. Each rules-based approach varies in predictive performance based on body part, as seen in Fig. 6. VCA generally outperforms JVSA except for duels involving the feet. NCA's underperformance relative to CA and VCA demonstrates the value of the decision frame, and VCA's superiority over CA confirms the utility of richer player representations.

3.3 Multimodal approach results

An aggregate accuracy of 80% was observed for MVC across the entire dataset. For multimodal ML-based methods, the distribution of test performances is shown in Table 3. For comparison, VCA performance on the same bootstrapped train/test splits is reported in the final (italicized) row.

Using the VCA 82% accuracy rate as a baseline, some multimodal methods (MC and BSMC) outperform the rule-based approach in expectation, but the difference is not statistically significant given the values of σ_{test}. In addition to aggregate accuracy, we continue the discussion of body-part-specific performance, with mean classification accuracies shown in Fig. 7. The overlapping distributions indicate that no single method consistently outperforms the others across all body parts.

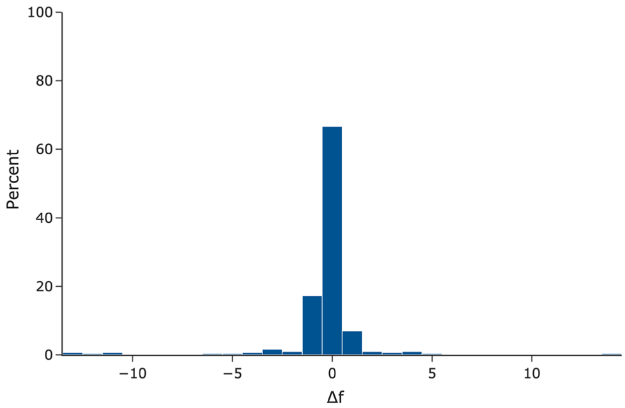


Fig. 4 Histogram of the differences between instrumented ball touch frame and the predicted decision frame; about 91% of decision frames occur within 1 frame of the instrumented football's reported touch frame

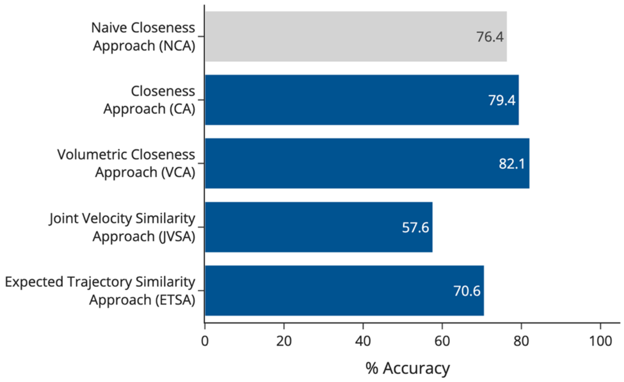


Fig. 5 Bar plot summarizing the classification accuracies of each rules-based approach using predicted decision frames

4 Discussion

In this paper, we demonstrated the ability of skeletal tracking data to aid out-of-bounds possession decisions. We began by defining heuristics that sufficiently isolated frames of interest for the duels. Next, distance-dependent rules-based approaches were able to predict decisions effectively, and we observed improvements as the closeness metric became more sophisticated. In addition, performance varied by body part. VCA adjudicated head duels with near 90% accuracy, but duels involving the legs or feet were around 65–75%. For these scenarios, JVSA outperforms VCA, likely due to noise in the tracking data under occlusion, making distance-based metrics less reliable. The multimodal approaches attempted to address this challenge by leveraging distance-independent features and supervised learning. However, since the last touch problem is, in principle, solved by VCA given perfect tracking of player bodies and the ball, and because the two approaches yield statistically indistinguishable results, we argue that VCA is the preferred method. Moreover, VCA and CA exhibit the highest importance in training, and VCA performance should improve naturally as tracking data quality improves.

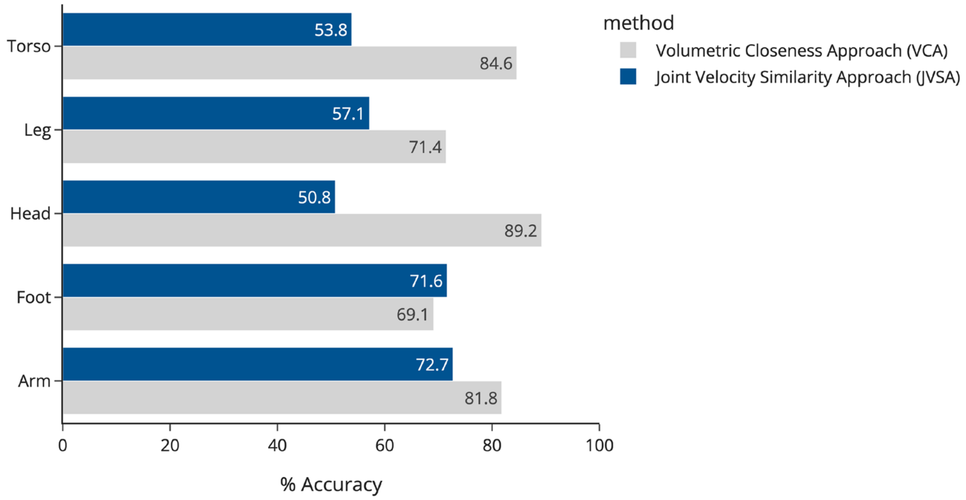
Table 3 Summary of MCCV Results for ML-based Multimodal Approaches with Rules-based VCA Comparison

Approach	μ_{train}	σ_{train}	μ_{test}	σ_{test}
Stacking Classifier (SC)	83.1	1.67	81.4	2.97
Margin Classifier (MC)	84.2	1.81	82.5	2.92
Body-segmented Stacking Classifier (BSSC)	84.6	1.37	81.4	3.22
Body-segmented Margin Classifier (BSMC)	86.4	1.65	82.5	3.17
Volumetric Closeness Approach (VCA)	*82.2*	*1.59*	*82.0*	*2.93*

We note that metric computation for rules-based approaches takes 1–2 s per duel, and inference runs in near-real-time, producing immediate predictions for individual duels across both rules-based and multimodal approaches. These findings suggest that SALT is fast enough for potential implementation during match play. Furthermore, speed improvements are likely achievable through parallelization, though this was not utilized in the current study.

The diphase design of SALT supports the need for adaptable decision support systems that are globally accessible. High definition multi-camera optical tracking systems could be replaced by more accessible broadcast tracking [31, 32]. The touch probability model could be substituted with precise touch data from connected balls, which has already been implemented in major international competitions [33–35]. Ultimately, SALT represents an advance in technology-assisted officiating that was previously infeasible with center-of-mass tracking, showcasing a potentially useful application of skeletal tracking.

Fig. 6 Comparison of JVSA and VCA classification accuracy by annotated body part



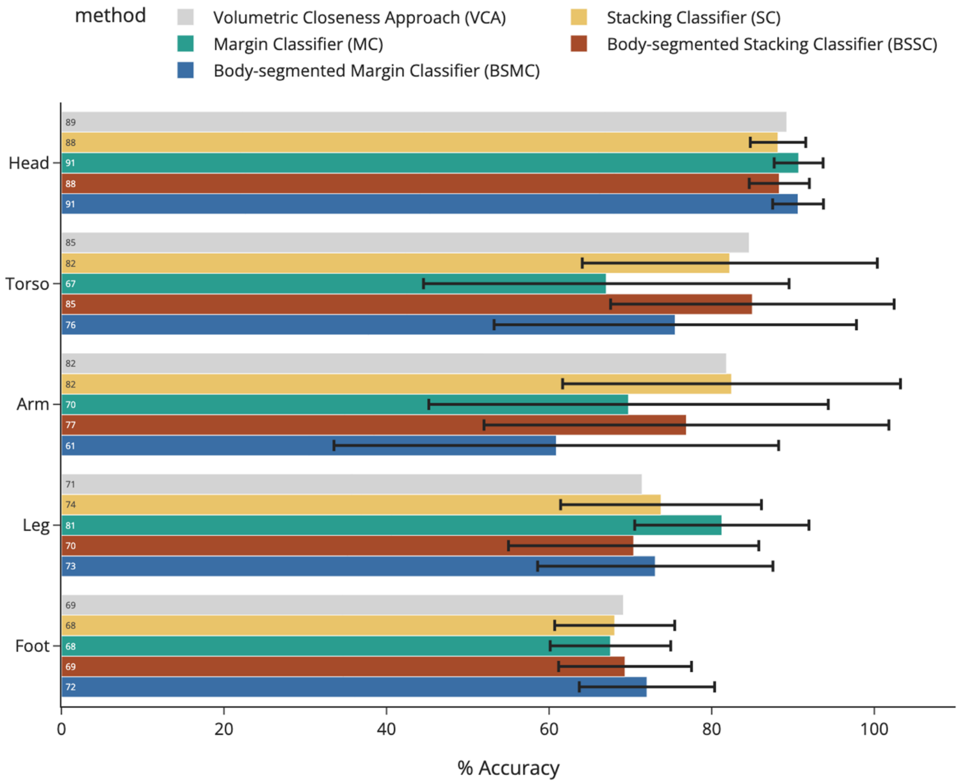


Fig. 7 Mean test accuracies for all multimodal approaches, segmented by body part, with rule-based VCA for comparison (light gray). Error bars denote standard deviations (SD) in classification accuracy across MCCV splits. SD bars all overlap within each body part

There are still many avenues for improvement for SALT, several of which are motivated by the limitations of this study. First, expanded testing on a diverse and representative dataset of duels, especially non-aerial duels, is needed to enable a more robust evaluation of performance, particularly given both the observed accuracy disparity between aerial and ground duels and the omission of non-visible duels during annotation, which could bias reported model performance. Second, pursuing improvements in the quality of the tracking data will benefit SALT in parallel, particularly under the preferred VCA method, since better tracking under occluded conditions and precise estimation of extremities will enhance the integrity of the system. Avatar or mesh representations can produce more accurate player models than standard volumetric templates [36]. Frame rates beyond 50 Hz will yield greater precision in the decision frame, as the contact time between the ball and the player is less than 0.01 s [37, 38]. The validation of labeled touch data will facilitate the training of the models for a more robust detection of the decision frames with better-balanced precision and recall. Finally, more work on measuring human accuracy on OOB decisions, which is currently lacking, is needed to understand SALT's true utility.

5 Conclusions

In this paper, we proposed several rules-based and multimodal approaches for Semi-automated Last Touch detection (SALT) in football using optical tracking data. Using touch data from an instrumented football to train a touch probability model, we were able to identify the frame of the last touch before the ball goes out-of-bounds. Then, by formulating skeletal and volumetric representations of players, measuring the relative movement of the ball and joints, and leveraging intuition on elastic collisions, we created four rules-based approaches to predict who touched the ball last in duel scenarios. A rules-based volumetric closeness approach achieved 82.1% accuracy in a criterion last-touch dataset, competitive against more sophisticated machine learning methods. We argue that this volumetric closeness methodology provides the most accurate representation of the physical system without the risk of overfitting our sample and may generalize better across diverse scenarios. Additionally, as tracking data continues to improve in quality and representativeness, this approach will become increasingly effective without requiring additional development. Ultimately, this work indicates that tracking data has the potential to effectively support possession decisions after out-of-bounds events and provides a baseline for future research.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12283-025-00518-3.

Acknowledgements The authors would like to thank FIFA and its partnering technology providers for provision of data and guidance, including Christian Hülsemeyer and Daniel Linke of KINEXON Sports. In addition, the authors are thankful for the expertise and assistance of Ana de Souza, Christina Chase, Ferran Vidal-Codina, and Ramzi Bensaid.

Author contributions KM collected the dataset used, including performing manual annotation. HW and AH performed the bulk of technical work and analysis, developing and implementing the methodologies used. HW conducted the analyses using Python, prepared the figures, and drafted the initial and revised manuscript text (for both revisions) with assistance from all authors. JB and SR guided the direction of the research design and types of research objectives to be targeted. All authors reviewed the manuscript multiple times and revised it critically for important intellectual content.

Funding 'Open Access funding provided by the MIT Libraries'. This research was funded by FIFA through the MIT Pro Sports Consortium. Two co-authors serve as Guest Editors for the Topical Collection II on Football Research in Sports Engineering. They had no role in the blind peer review process of this paper. All data subjects were informed ahead of competition that "Optical player tracking data, including limb-tracking data, will be collected, and used for officiating, performance analysis, research, and development purposes" and all relevant COUHES guidelines were adhered to.

Data Availability The data used for this study was collected by FIFA and collaborating data providers. Due to media and data rights, the datasets are not publicly available, but can be requested by contacting the authors of this article.

Declarations

Competing interests The authors declare no competing interests.

References

1. Mallo J, Frutos PG, Juárez D, Navarro E (2012) Effect of positioning on the accuracy of decision making of association football top-class referees and assistant referees during competitive matches. J Sports Sci 30(13):1437–1445
2. Spitz J, Put K, Wagemans J, Williams AM, Helsen WF (2016) Visual search behaviors of association football referees during assessment of foul play situations. Cogn Res Principles Implications 1(1):12
3. Fuller CW, Junge A, Dvorak J (2004) An assessment of football referees' decisions in incidents leading to player injuries''. Am J Sports Med 32(1):17–22
4. Dicks M, O'Hare D, Button C, Mascarenhas DR (2009) Physical performance and decision making in association football referees: a naturalistic study. TOSSJ 2(1):1–9
5. Vater C, Schnyder U, Müller D (2024) That was a foul! How viewing angles, viewing distances, and visualization methods influence football referees' decision-making. Ger J Exerc Sport Res 54(3):476–485
6. Wang H, Zhang C, Ji Z, Li X, Wang L (2024) Faster, more accurate, more confident? An exploratory experiment on soccer referees' yellow card decision-making. Front Psychol. https://doi.org/10.3389/fpsyg.2024.1415170
7. Spitz J, Wagemans J, Memmert D, Williams AM, Helsen WF (2021) Video assistant referees (VAR): the impact of technology on decision making in association football referees. J Sports Sci 39(2):147–153
8. Li M, Wang X, Zhang S (2024) The effect of video assistant referee (VAR) on match performance in elite football: a systematic review with meta-analysis. J Sports Eng Technol Proc Inst Mech Eng Part P. https://doi.org/10.1177/17543371241254596
9. Editorial, "The Guardian view on VAR: A slower, longer and fairer game may not be what football fans want," The Guardian, (2023), ISSN: 0261-3077. Accessed 15 Nov 2024
10. Buckingham P "The Premier League: VAR is working but checks do take too long," The New York Times, ISSN: 0362-4331. Accessed 15 Nov 2024
11. Cioppa A, Giancola S, Deliege A et al. (2022) "Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3491–3502
12. Cioppa A, Deliège A, Giancola S, Ghanem B, Van Droogenbroeck M (2022) Scaling up soccernet with multi-view spatial localization and re-identification. Sci Data 9(1):355
13. Giancola S, Cioppa A, Georgieva J et al. (2023) "Towards active learning for action spotting in association football videos," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5098–5108
14. Held J, Cioppa A, Giancola S, Hamdi A, Ghanem B, Van Droogenbroeck M (2023) "Vars: Video assistant referee system for automated soccer decision making from multiple views," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5086–5097
15. Held J, Itani H, Cioppa A, Giancola S, Ghanem B, Van Droogenbroeck M (2024) "X-vars: Introducing explainability in football refereeing with multi-modal large language models," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3267–3279
16. Lago C (2009) The influence of match location, quality of opposition, and match status on possession strategies in professional association football. J Sports Sci 27(13):1463–1469
17. Casal CA, Maneiro R, Ardá T, Marí FJ, Losada JL (2017) Possession zone as a performance indicator in football. The game of the best teams. Front Psychol. https://doi.org/10.3389/fpsyg.2017.01176
18. Jones PD, James N, Mellalieu SD (2004) Possession as a performance indicator in soccer. Int J Perform Anal Sport 4(1):98–102
19. Clark A, Corsie M, Gouveia V, Nunes NA (2024) "Fifa world cup 2022 qatar corner kicks: an analysis on effectiveness and match context," Qeios

20. Casal CA, Maneiro R, Ardá T, Losada JL, Rial A (2015) Analysis of corner kick success in elite football. Int J Perform Anal Sport 15(2):430–451
21. Beare H, Stone JA (2019) Analysis of attacking corner kick strategies in the fa women's super league 2017/2018. Int J Perform Anal Sport 19(6):893–903
22. Player & Ball Tracking System. [Online]. Available: https://inside.fifa.com/technical/football-technology/resource-hub?id=810f07b9d0694f0c9e43b653cbc51031
23. KINEXON LPS (Live). [Online]. Available: https://inside.fifa.com/technical/football-technology/resource-hub?id=776fb2814f8643de8dec1d6fda96b982
24. Bushby KM, Cole T, Matthews JN, Goodship JA (1992) Centiles for adult head circumference. Arch Dis Child 67(10):1286–1287
25. Cavia MM, Moreno A, Fernández-Trabanco B, Carrillo C, Alonso-Torre SR (2019) Anthropometric characteristics and somatotype of professional soccer players by position. J Sports Med Ther 4(4):073–080
26. Law 2 - The Ball — IFAB, https://www.theifab.com/laws/latest/the-ball/#qualities-and-measurements. Accessed 9 Sept 2024
27. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
28. van Erp M, Vuurpijl L, Schomaker L (2002) "An overview and comparison of voting methods for pattern recognition," in Proceedings eighth international workshop on frontiers in handwriting recognition, pp. 195–200. DOI: https://doi.org/10.1109/IWFHR.2002.1030908. (visited on 11/08/2024)
29. Robertson S, Zendler J, De Mey K et al. (2023) "Development of a sports technology quality framework," ISSN: 0264-0414
30. Sagi O, Rokach L (2018) Ensemble learning: a survey. WIREs Data Min Knowl Discovery 8(4):e1249
31. Yang Y, Li D (2017) Robust player detection and tracking in broadcast soccer video based on enhanced particle filter. J Vis Commun Image Represent 46:81–94
32. Yang C, Yang M, Li H et al (2024) A survey on soccer player detection and tracking with videos. Vis Comput. https://doi.org/10.1007/s00371-024-03367-6
33. Adidas reveals the first FIFA World Cup^{TM} official match ball featuring connected ball technology, https://news.adidas.com/football/adidas-reveals-the-first-fifa-world-cup--official-match-ball-featuring-connected-ball-technology/s/cccb7187-a67c-4166-b57d-2b28f1d36fa0, Jul. 2022. Accessed 22 Aug 2024
34. Brugts and Le Garrec top 'Connected Ball Technology' leader board after group stage, https://inside.fifa.com/technical/news/origin1904-p.cxm.fifa.com/brugts-and-le-garrec-top-connected-ball-technology-leader-board-after-group. Accessed 12 Feb 2024
35. Euro 2024: What is snickometer technology? How does snicko work? — DAZN News US, https://www.dazn.com/en-US/news/soccer/euro-2024-what-is-snickometer-technology-how-does-snicko-work/16oifal9vx87l102zxnl2vt2v1, Jun. 2024. Accessed 12 Feb 2024
36. Guo C, Jiang T, Chen X, Song J, Hilliges O (2023) Vid2Avatar: 3D Avatar Reconstruction from Videos in the Wild via Self-supervised Scene Decomposition, DOI: https://doi.org/10.48550/arXiv.2302.11566. arXiv:2302.11566. Accessed 12 Feb 2024
37. Nunome H, Lake M, Georgakis A, Stergioulas LK (2006) Impact phase kinematics of instep kicking in soccer. J Sports Sci 24(1):11–22
38. Lees A, Asai T, Andersen TB, Nunome H, Sterzing T (2010) The biomechanics of kicking in soccer: a review. J Sports Sci 28(8):805–817

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.