

A Framework for Automated Player Identification and Positioning Using Low-Cost Hardware in the Soccer Domain

Alexandre Cardoso Feitosa, Isaac Jesus da Silva, Danilo Hernani Perico

Electrical Engineering Department, Centro Universitário FEI, São Bernardo do Campo, SP, Brazil

Abstract

Soccer has reached a high technical, physical, and tactical level, making data use increasingly common among major clubs. Advances in extracting and processing players' data have encouraged more teams to explore this information. However, acquiring specialized data remains a challenge due to its complexity and high costs, which limits analysis and research. Publicly available data often include basic information like match results and player lineups, while commercial data, generated manually by individuals watching games, lack consistency. Positional data, crucial for advanced analysis, are typically obtained through expensive wearable GPS devices, limiting access to major soccer clubs. This research aims to propose a low-cost computer vision framework, named Advanced Player Identification and Positioning System, that is useful for extracting player positional data from television broadcast images. This approach enables the creation of advanced datasets, from which essential information for soccer can be extracted, such as tactical formation, for example. The proposed framework was divided into four steps: detection, tracking, identification, and positioning. Experiments were conducted across all stages of the APIPS system using the SoccerNet-GSR dataset, with its manual annotations serving as ground truth. The results indicate that player identification can be improved by the proposed temporal tracking strategy. Furthermore, except for a few outlier cases, the final player positioning error was below 5 meters in 91% of the evaluated instances.

KEYWORDS: DATA ACQUISITION, MACHINE VISION, OBJECT TRACKING, SPORTS

Introduction

The evolution of sports over the years has been influenced by technological and scientific advancements (Frevel et al., 2022). In the context of soccer, during its early stages of development and popularization, the main focus of clubs worldwide was on improving players' technical and physical abilities. However, scientific advances have made it possible to analyze the game in a broader and more systematic way, incorporating, for instance, data analysis for strategic decision-making (Barnes et al., 2014; Sarmento et al., 2018).

One case in which data contributed to sporting success occurred in the quarterfinals of the 2006 FIFA World Cup, when German goalkeeper Jens Lehmann used information about his opponents' penalty-kick preferences to decide which side to dive. Lehmann saved two of the four shots, helping qualify Germany for the semi-finals (Perl, Grunz, & Memmert, 2013). Another case took place during the 2010 FIFA World Cup, when analysts identified a high likelihood of goals being conceded by England when their opponents attacked using long balls. Germany then used this information to score one of the four goals in the 4-1 victory over the English team (Perl et al., 2013). These are examples that illustrate how performance analysis in soccer can contribute to the preparation phase for games.

Despite increasing interest in soccer analytics, the academic community still lacks quality public datasets (Pappalardo et al., 2019). The few studies that have been carried out are typically done when there is a partnership with a club or federation, or when researchers choose to manually record data, which demands substantial time and effort. The advancement of soccer analytics depends on access to objective, high-quality data. Unlike traditional assessments based on subjective observation, data-driven methods quantify actions and tactical behaviors with precision. Metrics such as passing networks, space control, and positioning patterns require detailed spatiotemporal information. Without reliable datasets capturing match events and player movements, analytical approaches become limited. Thus, expanding the availability of structured and accurate data is essential to producing meaningful insights for competitive decision-making (Majeed et al., 2025; Lolli et al., 2025).

Gudmundsson & Horton (2017) highlighted that two main types of spatiotemporal data are usually collected in team sports: player or ball trajectories and event logs describing actions such as passes or shots. Each dataset enables important analyses on its own, for example, movement can reveal team formations, while event logs indicate contextual factors such as ball possession. These data can be automatically captured by devices like GPS vests or cameras.

The evolution of wearable technologies, such as GPS vests, has significantly increased the accuracy of player positional data on the field (Rampinini et al., 2015). This advancement enabled the analysis of new metrics, such as distance covered and positioning patterns, which can support a better understanding of opponents' tactical formations, playing styles, and players' behavior on the field. However, this is an expensive technology for most soccer clubs. As a result, many organizations lack the financial resources needed to acquire such systems, and although there are cases in which these data have contributed to improved performance, the topic remains underexplored and its practical application is still limited. Wright et al. (2012) reported that only 2% of coaches select key performance indicators based on academic literature. This gap can be largely attributed to the complexity of the data and the high costs involved.

Consequently, computer vision has become one of the most significant technological drivers in soccer analytics. Automatic video analysis enables, for instance, the tracking of players and the ball on the field, as well as the measurement and review of game actions such as shots on goal, missed passes, and ball possession to be performed directly from match footage. Thus, computer vision allows the acquisition of trajectory and action-related data, ultimately making the entire

process faster and more reliable (Majeed et al., 2025; Giancola et al., 2018).

Both wearable technologies and computer vision systems have contributed to the significant growth in the volume of data, creating countless research possibilities in sports analytics. With advancements in big data manipulation techniques (Krishnan, 2013), numerous soccer performance metrics can be extracted and computed with greater accuracy and efficiency (Memmert & Rein, 2018; Morra et al., 2020; Manescu, 2025; Majeed et al., 2025).

Focusing specifically on computer vision-based player and ball tracking throughout a match, it should be noted that few works have been successful in popularizing the use of their proposed frameworks. In most cases, the main challenge arises during the occlusion of players' jersey numbers, as highlighted in Theiner et al. (2022), Diop et al. (2022), and Alhejaily et al. (2023). Despite notable advances in the field, no method has fully overcome the challenges of tracking and identifying players using broadcast-only footage. Moreover, it is even more difficult to find research that successfully extracts player position data from broadcast videos using computer vision under constrained computational resources.

Therefore, the objective of this work is to propose, implement, and analyze the Advanced Player Identification and Positioning System (APIPS) framework, designed to obtain player positional data from television broadcast images, using low-cost computer resources. In addition, the project also has the following secondary objectives:

- Propose, implement and analyze the Temporal Tracking Identifier (TTI) algorithm, designed to identify players based on their jersey numbers across consecutive frames, even in situations of occlusion or changes in appearance.
- Ensure project execution using limited computing resources, such as an NVIDIA GeForce GTX 1060 6GB GPU, demonstrating the viability of affordable analysis systems on low-cost hardware.

Background and Related Works

Several studies have shown that the application of machine learning techniques can be a determining factor in the success of sports clubs. In order to organize them, this section was divided into three: the first covers studies that focus on the use of pre-existing data to obtain benefits in the sports field; the second encompasses works that aim to generate or process raw data, which can later be used for analysis; the third brings together works that have as their main focus the tracking of multiple objects. The graph in Figure 1 shows the number of works found by year of publication, demonstrating how the topic has grown in recent years, specifically since 2018, the year of the World Cup held in Russia, in which the Video Assistant Referee (VAR) was used for the first time in an official FIFA competition. Since then, several competitions have adopted this and other technologies, such as automatic offside detection.

Use of Pre-existing Data

All the works studied in this section used datasets generated by companies specialized in capturing, preprocessing, and selling data; therefore, the datasets are not publicly available.

The work presented by Cortez et al. (2022) applied machine learning to identify, based on the physiological data of the players, which are best suited to be part of the starting lineup in a match. The methodology developed resulted in an index, called the lineup preparation index, for the following player positions: fullbacks, midfielders, and wingers. For the other positions, the results were not satisfactory.

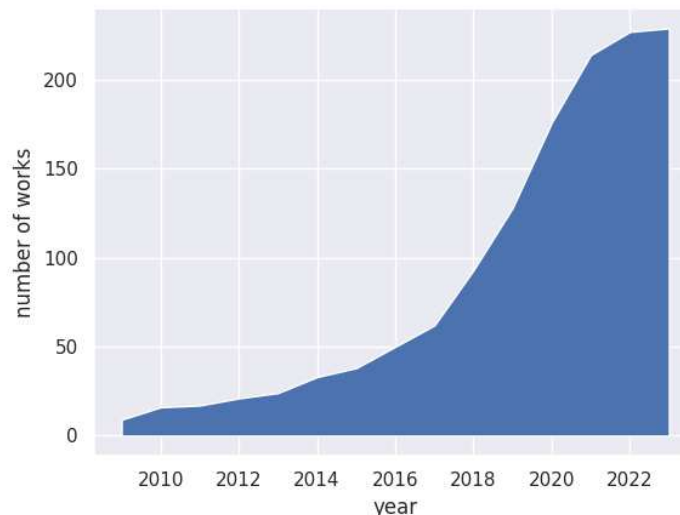


Figure 1 - Quantitative analysis of studies carried out in the focus area over the last decade: computer vision being used in soccer

Positional data was also explored by Buyrukoglu & Savaş (2023) to classify the players' positions on the field. The authors adopted techniques such as Random Forest, Gradient Boosting, and Deep Neural Networks to obtain accuracy of 83.9%. In another study, machine learning techniques were used to propose a new method to estimate the financial value of players in the transfer market (Behravan & Razavi, 2021).

Another subtopic explored with machine learning is the analysis of individual player performance. Manish et al. (2021) designed a system with several deep learning models, one for each position. In this way, it was possible to increase assertiveness, as it was noted that certain variables (correct passes, complete tackles, among others) behave differently for each of the positions. On the other hand, Jamil et al. (2021) used machine learning algorithms to classify elite and sub-elite goalkeepers in professional men's soccer. The results discovered in this study suggest that skill with the feet and not necessarily with the hands is what distinguishes elite goalkeepers from sub-elite goalkeepers.

Despite the most diverse fields of study that involve machine learning and soccer, predicting results remains the topic of most research. The work presented by Vashist et al. (2021) adopted a set of machine learning techniques to predict the outcome of Premier League matches (English championship) with an accuracy of 75%. Similarly, Rahman (2020) used deep neural networks to obtain an accuracy of 63.3% on the results of the 2018 FIFA World Cup. In contrast, Yang (2021) applied machine learning to predict the outcome of matches but does so based on the performance of the players and not just on previous results. With this methodology, the author achieved an accuracy of 79.6%.

However, something that is still little explored in the articles analyzed is the hypothesis testing considering a mixed set of information. Most studies start from historical results, and in some cases, as in Cortez et al. (2021), the physiological variables of the players are considered. Nevertheless, no studies were found in which these and other variables (such as weather conditions, style of play, positional data) are considered in a combined way.

Object Tracking

In the work presented by Ciaparrone et al. (2020) object tracking is defined as the process of analyzing the position, scale, state, and other information of the target in consecutive video frames, predicting the behavior and trajectory of movement to continuously track the specific

target.

Zhang et al. (2022) proposed ByteTrack, a multi-object tracker capable of outperforming the main tracking algorithms, mainly due to the addition of an algorithm that reuses low-scoring detection boxes to associate them with high-scoring detection boxes. This makes the tracker more resistant to occlusion, motion blur, and changes in the size of the tracked objects. In addition, this is one of the few articles that share the original source code, which makes it easy to take advantage of what has been done.

SportsMOT (Cui et al., 2023) was presented as a multi-frame object tracking dataset from different sports, such as Volleyball, Basketball, and Soccer. In addition, the authors also propose an algorithm for multi-object tracking called MixSort, which makes use of MixFormer (Cui, Jiang, Wu, & Wang, 2024). In Cioppa et al. (2022), another dataset with 12 complete soccer matches is presented, as well as a general review of player tracking methods in videos that contributes to understanding the level of maturity that the subject has reached in recent years.

Nevertheless, one of the most widely used algorithms for sports tracking is the Simple Online and Realtime Tracking with a Deep Association Metric (DeepSORT) (Wojke, Bewley, & Paulus, 2017). This algorithm extends the original SORT approach (Bewley, Ge, Ott, Ramos, & Upcroft, 2016) by incorporating deep learning to enhance the association of detections across frames, even in challenging scenarios involving occlusions or abrupt object motion. DeepSORT combines Kalman filter-based tracking with deep learning to achieve more robust performance. Furthermore, DeepSORT has low computational costs when compared to ByteTrack, for example.

Scott et al. (2024) introduced TeamTrack, a benchmark dataset for multi-object tracking (MOT) in team sports, including soccer, basketball, and handball. The dataset comprises over 4 million annotated bounding boxes in high-resolution videos captured from side and top views, providing comprehensive coverage of the playing field. The authors evaluated existing MOT methods on TeamTrack, highlighting challenges in tracking players with similar appearances under high occlusion and complex motion. Experiments evaluated the performance of the proposed algorithm on the TeamTrack dataset, demonstrating its behavior in multi-object tracking tasks within sports contexts.

Yang et al. (2025) provided a comprehensive survey on soccer player detection and tracking in videos, reviewing state-of-the-art methods such as DeepSORT and TrackFormer. The survey details preprocessing techniques, including background subtraction and perspective transformation, as well as postprocessing strategies for mapping players onto a 2D field representation. It highlights challenges in real-world scenarios, such as occlusions, visually similar players, and complex motion patterns, and provides a comparative evaluation of existing approaches. While the survey emphasizes the scarcity of comprehensive datasets for evaluating multi-object tracking (MOT) methods, Scott et al. (2024) addressed this gap by introducing TeamTrack, a benchmark dataset comprising over 4 million annotated bounding boxes in high-resolution videos from soccer, basketball, and handball.

Vision-Based Localization and Tracking

The second stage of the literature review was to analyze works that aim to obtain data in sports and understand which methods can be used to allow the creation of a dataset containing positional information in the soccer domain. In Theiner et al. (2022), a complete pipeline is proposed that covers everything from camera detection and frame selection stages to the estimation of the player's position. However, the authors conclude that to enable player-based analysis throughout a game, an additional module focused on the location and re-identification of players is necessary.

A player location system, focusing on the recognition stage, was also proposed by Diop et al. (2022). The authors used the YOLO (Redmon & Farhadi, 2018) object detection algorithm to detect players on the field and subsequently used the Dlib toolkit (King, 2009) to perform facial recognition of players in close-up frames. For frames with a wide view of the field, a Convolutional Neural Network was employed with the public MNIST database to recognize the player's jersey number. However, for frames in which one or more players had their respective jersey numbers hidden, there is no recognition.

Alhejaily et al. (2023) proposed a player localization system composed of the stages of field detection, player detection, team assignment, and player identification by jersey number. The player detection stage was performed using the YOLO model (Redmon & Farhadi, 2018), and the team assignment used a special type of convolutional neural network, called Convolutional Autoencoders - CAEs. However, the player identification stage was performed using the jersey number as a reference, which serves as input for a Deep Neural Network, and, as in the previous work, in cases where this number cannot be seen, such as in a frame in which the player is facing the camera, there is no player identification. Finally, according to the authors, future work in the area could focus on tracking players, estimating their position on the field over time, and recognizing actions such as passes, shots, and fouls.

The work presented by Somers et al. (2024) formalizes the task of Game State Reconstruction and introduces SoccerNet-GSR, a dedicated dataset for soccer video analysis. SoccerNet-GSR comprises 200 video sequences, each 30 seconds long, annotated with 9.37-million-line points for pitch localization and camera calibration, as well as over 2.36 million player positions, including metadata such as role, team, and jersey number. However, the pipeline presented in this work needs a high-performance GPU to work properly; they used an NVIDIA A100 32GB.

Golovkin et al. (2025) proposed a vision-based framework for player tracking and identification using broadcast video. They also used the SoccerNet-GSR dataset and they address key challenges such as occlusion, camera viewpoint changes, and identity consistency across frames, offering solutions tailored to the limitations of single-camera setups. The approach enhances the extraction of positional data for sports analytics under real-world constraints. Experiments were conducted using an NVIDIA A100 GPU.

Majeed et al. (2025) proposed a real-time framework to analyze ball–player interactions in soccer videos, combining CSPDarknet53-based detection with a graph convolutional network (GCN) to model player–ball relationships. The method derives interaction metrics (e.g., proximity and inter-player distances) and estimates physical indicators such as distance covered and speed. The authors report approximately 91% object-detection accuracy, 90% performance for tracking and action recognition, and 92% for speed analysis, along with improvements over prior GCN-based approaches, suggesting robustness for extracting tactical insights in realistic video settings.

Considerations

All the reviewed articles are important for understanding the progress made in frame detection, field delimitation, and player recognition.

Studies such as Cortez et al. (2022) and Buyrukoglu & Savaş (2023) applied machine learning to player selection and positional classification, while predictive models by Vashist et al. (2021), Rahman (2020), and Yang (2021) focused on match outcomes. These works highlight the value of integrating diverse information sources but often rely on structured datasets, wearable sensors, or historical statistics, limiting real-time applicability.

In the area of positional data extraction, Theiner et al. (2022), Diop et al. (2022), and Alhejaily

et al. (2023) proposed detection and identification pipelines using YOLO and CNN-based jersey recognition. Although effective, these methods encounter challenges when jersey numbers are occluded or not visible, and continuous tracking across the full field remains limited. Multi-object tracking approaches, including DeepSORT (Wojke, Bewley, & Paulus, 2017), ByteTrack (Zhang et al., 2022), SportsMOT (Cui et al., 2023), and TeamTrack (Scott et al., 2024), improve robustness to occlusion and motion blur but depend on annotated datasets, wearable devices, or high-performance hardware. Yang et al. (2025) provides a survey of these methods, emphasizing the scarcity of large-scale datasets suitable for broadcast video analysis.

However, there is a lack of publicly available studies that can effectively track players during a match and subsequently determine their positions on the field, especially using limited computational resources.

The present study extends these foundations by proposing the Advanced Player Identification and Positioning System, which operates directly on television broadcast images. The framework integrates detection, tracking, identification, and positioning, allowing continuous player localization even under occlusions or when jersey numbers are not visible. Temporal tracking enhances identification accuracy, and the final positioning achieves errors below 5 meters in over 90% of cases. In addition, the system enables the creation of advanced datasets that support tactical analysis, such as formation reconstruction, providing an automated solution for positional data extraction in professional soccer.

Advanced Player Identification and Positioning System

To build the Advanced Player Identification and Positioning System (APIPS), this work uses a four-step methodology, as illustrated in Figure 2. The application is exemplified in Figure 3.

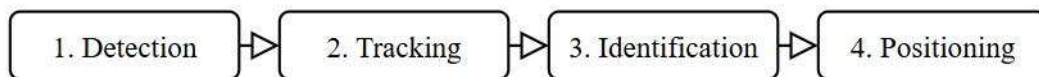


Figure 2 - Methodological approach to APIPS development

The video is received and each of its frames undergoes detection, in which the output is the bounding boxes, indicating in which regions there are people or balls. Then, the bounding boxes of the different frames are associated through tracking, which assigns a unique identifier to each of the objects throughout the frames. Once tracked, the team and number are identified using the information present on the player's shirt. Finally, the position is estimated for each of the bounding boxes in relation to the field.

Input Frames

The input stream consists of a sequence of frames obtained through a camera positioned in the middle of the field. Excerpts from broadcasts are often made publicly available. In this project, a public dataset was chosen: the SoccerNet-GSR¹ (Somers et al., 2024), which offers not only excerpts of matches recorded with single cameras and without cuts, but also a robust set of manual annotations related to the stages of detection, tracking, team identification, number and positional data of the players.

¹ <https://github.com/SoccerNet/sn-gamestate>

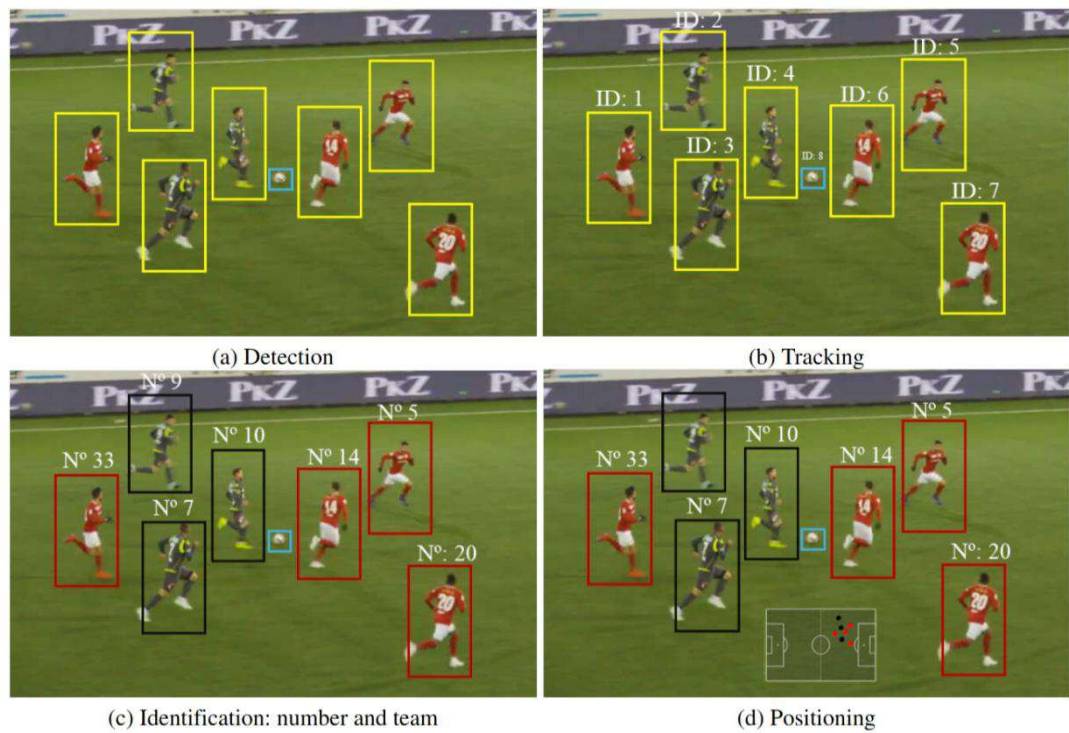


Figure 3 - Images manually created to illustrate APIPS flow.

Detection

The objective of this step is to detect, in each video frame, the largest number of players and other game elements (referees and ball), returning the coordinates and dimensions of the bounding boxes. To this end, the YOLO algorithm (Wang et al., 2024; Jocher & Qiu, 2024) is used, motivated by the good results presented in recent works, such as in Diop et al. (2022); and Alhejaily et al. (2023). Different versions (10 and 11) and sizes (L, M and S) of YOLO are tested, as well as input resolutions (1920, 1280 and 640 pixels) and confidence thresholds ranging from 0.1 to 0.8, in order to balance accuracy and processing time. As an output, the method returns, for each object found in each frame, a bounding box containing its respective coordinates and dimensions.

Tracking

The tracking stage aims to associate the bounding boxes detected throughout the video frames, assigning unique identifiers to each object to enable its tracking frame by frame. To this end, the DeepSORT algorithm is used, which presents high computational efficiency and good feature extraction capacity, in addition to having already achieved a MOTA (Multiple Object Tracking Accuracy) (Bernardin & Stiefelhagen, 2008) of 94.84% in tracking soccer players (Cioppa et al., 2022). The main parameters to be evaluated are:

- *max_age* (1, 3, 5, 10, 25, 75, 300): number of frames in which an object may not appear before being discarded.
- *n_init* (1, 3, 5, 10): minimum number of consecutive frames to start a new tracking.
- *max_cosine_distance* (0.1 to 0.7, steps of 0.2): similarity threshold for embeddings association.

Bounding boxes are associated with unique identifiers, allowing each player to be tracked across frames and supporting subsequent identification and positioning stages.

Identification

The identification stage aims to first assign the corresponding team to each tracked bounding box and then recognize the number on each player's jersey. To distinguish teams and referees, the color of uniforms is analyzed using histograms extracted from each box, calculating the average of the values. The K-Means clustering algorithm (MacQueen, 1967; Lloyd, 1982) is then applied to separate these averages into three groups: Team A, Team B, and the referees. This approach, in addition to being unsupervised, tends to be efficient due to the striking color differences in the uniforms.

Identifying a player's number requires additional care, as the number is not always visible since it is printed on the back of the jersey. As discussed in Alhejaily et al. (2023), the physical similarity between athletes and the limited quality of cameras makes it difficult to use other characteristics, such as face or hairstyle. To estimate the number of shirts in each box, a model based on Optical Character Recognition (OCR) is built. The implementation uses the EasyOCR Library (Shi, Bai, & Yao, 2017; JaideAI, 2020), which offers several adjustable parameters. Some of them are kept fixed, such as *detail=1* and *allowlist='0123456789'*, to limit detection exclusively to digits and obtain a corresponding confidence measure for the prediction. Other parameters are explored for optimization, such as the variation of input resolution (64 to 640 pixels), decoder modes (*beamsearch* and *greedy*), and different threshold ranges (*text_threshold*, *low_text*, *link_threshold*, *width_ths*, *height_ths*) between 0.3 and 0.8. This procedure seeks to balance detection accuracy with computational efficiency.

Ultimately, each bounding box is associated with a team and a number, and a confidence level is obtained for the detected number. This enables players to be identified and differentiated on the field with a high degree of accuracy, even in scenarios where the number on their shirts is not fully visible.

Temporal Tracking Identifier

In many frames, the player's number is not visible, which limits identification to analyzing only a single frame. To overcome this problem, the Temporal Tracking Identifier (TTI) is proposed, which explores information throughout the entire video. If the number is visible in any frame, it is identified and spreads to others in which it does not appear clearly.

TTI considers information originating from the detection, tracking, and identification stages (positions of the bounding boxes, tracking identifier, team, and number). Thus, the method assigns a unique number to each player. Although conventional methods are limited to frames in which the shirt is perfectly visible, TTI spreads the recognized number throughout the temporal sequence, overcoming situations in which the athlete is seen head-on. To infer the most likely number, TTI supports multiple heuristic formulations. In this study, we investigate the following heuristics:

- **Highest Frequency:** chooses the most predicted number. For example, for a player tracked on 5 frames who had predictions 2, 21, 21, 2, and 2, TTI propagates the number 2. In case of a tie, the algorithm propagates the number with the highest average confidence.
- **Highest Confidence:** uses only the prediction with the highest confidence. For example, for a player tracked on 5 frames who had predictions and confidences of (2; 0.7), (21; 0.8), (21; 0.1), (2; 0.6), and (2; 0.7), TTI chooses to propagate the number 21 because it had the highest confidence (0.8). In case of a tie, the algorithm would use the number with the highest frequency.
- **Highest Average Confidence:** propagates the number with the highest average

confidence. For example, for a player tracked over 5 frames who had predictions and confidences of (2; 0.7), (21; 0.8), (21; 0.1), (2; 0.6), and (2; 0.7), TTI chooses the number 2 because it has the highest average confidence among the predictions: $(0.7 + 0.6 + 0.7)/3$ against $(0.8 + 0.1)/2$.

- **Combined Digit Highlighting:** it assumes that a player may have a number with more than one digit, such as 21 or 34, but in some frames the algorithm may only see a single digit (2 or 3, for example). To deal with this scenario, this approach checks whether there is any multi-digit number that has been predicted with good average confidence (above a defined threshold) on at least two occasions. If found, this multi-digit number is chosen to be propagated.

TTI is designed to improve identification even in unfavorable scenarios, increasing the number of correctly recognized players. Furthermore, by adding temporal evidence, the method can correct inaccurate point predictions, favoring the alternative with greater global confidence.

Positioning

After detecting, tracking, and identifying each player, referee, and ball throughout the frames, it becomes necessary to estimate the position of these objects in relation to the playing field. To do this, the camera orientation is defined in each frame, precisely locating the filmed region, and from reference points (such as penalty area markings), the exact position of the bounding boxes is calculated. These positions are projected onto a 2D matrix that represents, over time, the entire field. To obtain the camera orientation and perform the projections, the TVCalib framework (Theiner & Ewerth, 2023) is used, which is responsible for calculating the homography matrix, which relates the camera and field coordinate systems. This matrix converts the positions observed in the images to the spatial field domain, correcting distortions, and unifying the information from different frames into a continuous and coherent view.

Next, the base of the bounding boxes (the players' feet) is identified and the homography matrix is applied to project these coordinates onto the field plane. After projection, the values are normalized to provide Cartesian coordinates (X, Y). Thus, APIPS generates positional data for players, referees, and the ball, with their respective team identifiers and numbers. Finally, it is possible to display this information in a 2D field that represents the arrangement of objects.

Experiments and Results

This section details the experiments that were applied in each stage of the APIPS system, with the aim of validating the quality of the constructed process. In all cases, the manual annotations provided by the SoccerNet-GSR database were used as a reference. The experimental evaluation was conducted on five video sequences, each 30 seconds long and recorded at 25 frames per second, resulting in 750 frames per video (i.e., $30 \times 25 = 750$).

Object Detection

This experiment evaluated YOLO's object-detection performance using the F1-score. For each IoU (Intersection over Union) threshold $t \in \{0.1, 0.2, \dots, 0.8\}$, we computed the corresponding true positives (TP), false positives (FP), and false negatives (FN) by matching each predicted bounding box to the ground-truth and counting it as a TP when $\text{IoU} \geq t$; unmatched ground-truth boxes were counted as FNs and unmatched predictions as FPs. The F1-score was then computed at each threshold, and the final reported score corresponds to the average F1 across all thresholds. Different configurations were tested, varying version, size, resolution, and confidence threshold, as described in the last section. The F1-score results for versions 10 and 11 indicate that

YOLOv10 slightly outperformed YOLOv11 (76.8% vs. 75.6%), at the cost of marginally slower inference. With respect to model size, the S, M, and L variants achieved F1-scores of 74.0%, 75.9%, and 78.5%, respectively, where each value is averaged across the YOLOv10 and YOLOv11 results. Regarding input resolution, 1280 provided the best trade-off, reaching an F1-score of 80.0% (averaged across v10 and v11), outperforming 640 (72.3%) and 1920 (76.1%). Table 1 summarizes these comparative results. Among all tested settings, YOLOv10-L at 1280 resolution with a 50% threshold achieved the highest F1-score (83.8%) and was therefore selected for the subsequent stages of the proposed method. Figure 4 illustrates example detections made by this best-performing configuration.

To assess visual consistency, the real bounding boxes (via SoccerNet-GSR) were displayed next to the boxes detected by the best model, concluding that the detection was mostly satisfactory, although it oscillates in high-speed situations or occlusions. In such cases, distortion at the edge of the image and the presence of extra characters (e.g. referees and ball boys) can influence the result. Still, it is reinforced that the detection focused only on finding objects in the field, without differentiating their category, resulting in a number of detections greater than the total number of players.

Table 1 - F1-score Metrics per Model Configuration

Configuration	F1-score	Observations
YOLO v10	76.8%	Better than Version 11, but slightly slower inference
YOLO v11	75.6%	Faster inference
Size S	74.0%	Faster inference, lower F1
Size M	75.9%	Balanced
Size L	78.5%	Higher inference cost
Resolution 640	72.3%	Lower performance
Resolution 1280	80.0%	Best balance between F1 and inference
Resolution 1920	76.1%	Slower inference
Best Configuration	83.8%	Version 10, Size L, Resolution 1280, 50% threshold; best overall performance

Tracking over Time

This section presents the experiments carried out to evaluate the performance of the DeepSORT algorithm in the task of tracking multiple objects. Recognized metrics from the tracking literature were used, such as precision, recall, F1-score, and MOTA, using the manual annotations of SoccerNet-GSR as a reference (real IDs and coordinates of the boxes).

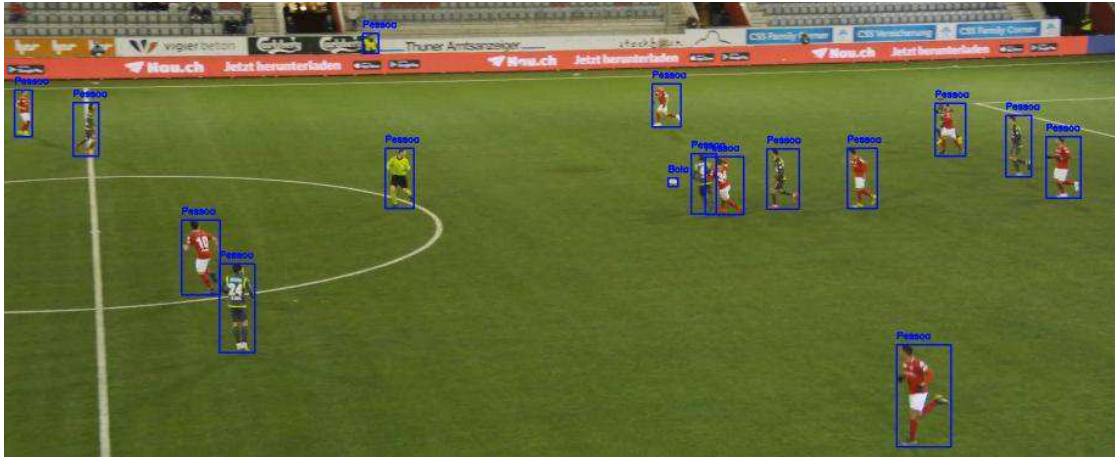


Figure 4 - Detection made by the best model: YOLO v10-L, resolution 1280 and threshold of 50%

Precision indicates the percentage of correctly tracked boxes in relation to the total tracked, while recall shows the proportion of tracked boxes in relation to the actual total of objects. The F1-score combines both, giving the same relevance to precision and recall. MOTA penalizes detection errors (FP, FN) and identifier changes, evaluating the overall quality of tracking.

The parameters *max_age* (1, 3, 5, 10, 25, 75 and 300), *n_init* (1, 3, 5 and 10) and *max_cosine_distance* (10%, 30%, 50%, 70%) were tested. Table 2 shows the best evaluated configurations. The best models are observed to have exceeded the F1-score of 91% and the MOTA of 81%, indicating a high level of tracking. The smaller *max_age* and *n_init* parameters produced superior results, since objects that suddenly disappear are not tracked for long intervals. Furthermore, the results also suggest that smaller values of *max_cosine_distance* are better, indicating that the embeddings used do not maintain good similarity in high-variation situations.

Table 2 - Results for the best models generated with different parameters of the DeepSORT algorithm

ID	<i>max_age</i>	<i>n_init</i>	<i>max_cosine</i>	Precision	Recall	F1-score	MOTA	Time (s)
1	1	1	0.1	94.2%	88.8%	91.4%	81.5%	65
2	1	1	0.3	94.2%	88.8%	91.4%	81.5%	66
3	3	1	0.3	92.2%	90.2%	91.2%	81.1%	71
4	3	1	0.1	92.2%	90.1%	91.1%	81.0%	67
5	3	3	0.3	92.7%	88.9%	90.8%	80.6%	67
6	3	3	0.1	92.6%	88.9%	90.7%	80.4%	67
7	1	3	0.1	94.5%	86.9%	90.6%	80.3%	65

Qualitative analysis showed that the greatest difficulty occurs in scenarios of partial occlusion (as exemplified in Figure 5) or temporary disappearance, although tracking performance was generally reliable.

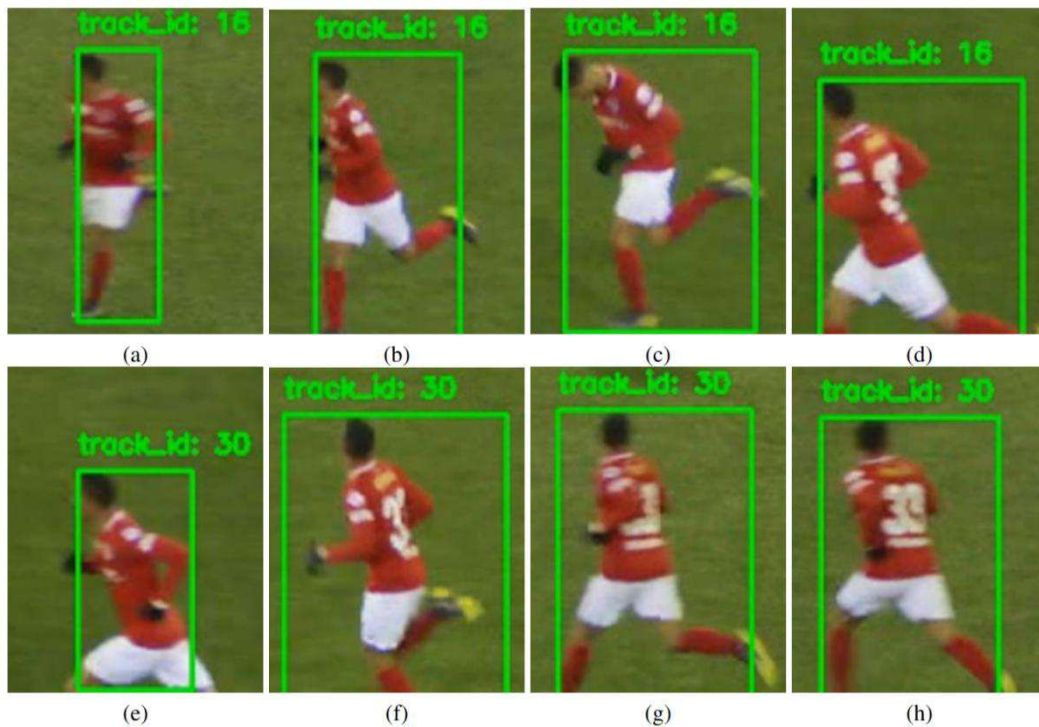


Figure 5 - Example of partial occlusion: player begins to leave the camera's field of view in the frame presented in (d). In (e) there is an exchange of the identifier

Player Identification

This section evaluates the performance of the identification step in two sub-steps: team identification and number identification. To classify each bounding box as Team A, Team B or referee, K-Means was applied, comparing the predicted labels to the manual annotations of SoccerNet-GSR. Since the clustering algorithm does not indicate the side of the field, the three labels were mapped to the real set according to the highest possible precision. Thus, it was possible to calculate precision, recall and F1-score. Table 3 presents the performance obtained.

Table 3 - Results of the proposed methodology for team classification

Class	Precision	Recall	F1-score
Team A	86.2%	97.5%	91.5%
Team B	93.5%	95.7%	94.6%
Referee Team	89.6%	53.7%	67.1%

Both teams had a high F1-score, which confirmed that the color of the uniform is a decisive factor for the grouping. The Refereeing Group obtained lower recall, since the goalkeeper and referee uniforms were sometimes confused by the model, which reduced the performance in relation to refereeing.

For the identification of numbers, two metrics were adopted: total accuracy (%), which computes accuracy in all boxes, and a conditional accuracy, which excludes boxes where the number is not visible. Several models were generated by varying parameters to maximize the percentage of total accuracy. In the end, the resolution of 512 proved to be the most advantageous. Table 4 shows the comparative results.

Table 4 - Results of the models with the best parameters, for different resolutions

Resolution	Total Accuracy	Conditional Accuracy (visible digit)	Time (s)
640	1.73%	5.07%	883
512	5.34%	15.73%	740
256	3.30%	20.49%	232
128	1.86%	26.22%	101
64	0.16%	8.77%	69

The total accuracy was low, since most of the boxes did not display any number. Furthermore, the final accuracy was compromised in frames where only one digit was visible or the player was partially hidden, as depicted in Figure 6. However, in situations of full visibility, the predictions were often correct. It can therefore be concluded that frame-by-frame identification is insufficient when the digits are hidden or distorted, reinforcing the need for strategies such as TTI, which take advantage of temporal information to propagate correct identification.

Therefore, the experiment considering TTI was conducted. It aimed to evaluate the gain in accuracy for player identification when using the TTI algorithm in images with a resolution of 512. As discussed previously, it is not feasible to analyze only one frame, since most of them display numbers that are invisible to the model. Thus, the value propagated by the TTI is adopted as the final prediction. As previously presented, four heuristics were tested.

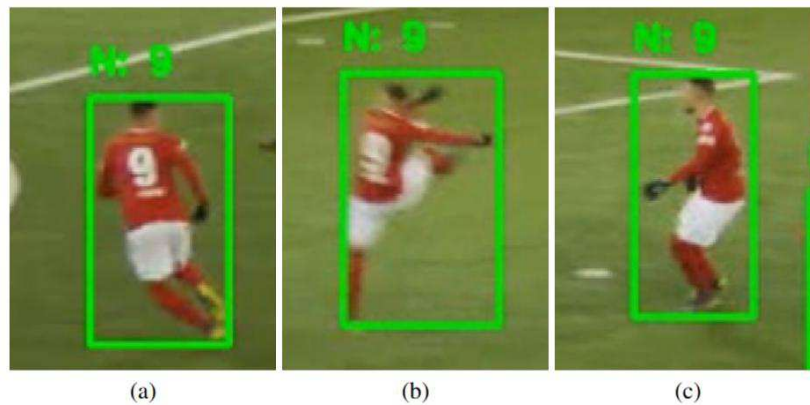


Figure 6 - Example of possible cases during the process of identifying the player: (a) only part of the number is visible – wrong identification; (b) both digits are visible – correct identification; (c) no digit is visible – no identification.

Table 5 summarizes the results. TTI increases the proportion of hits; TTI with Combined Digit Highlighting (Figure 7) achieved the best result, reaching 26.27%. This increase indicates that temporal propagation mitigates the problems faced when the number is partially observable in isolated frames. The lower hit rate achieved by TTI with Highest Frequency confirms the high rate of false positives in cases of partial occlusion.

Table 5 - Results obtained with (and without for reference) TTI considering the different cases of heuristics.

Heuristic	Total Accuracy
No TTI	5.34%
TTI – Highest Frequency	12.71%
TTI – Highest Confidence	18.19%
TTI – Highest Average Confidence	13.74%
TTI – Combined Digit Highlighting	26.27%

Despite these advances, approximately 75% of the boxes still lack a reliable prediction. Qualitative analysis indicates that the short duration of the videos (30 seconds) and the lateral position of the camera make it difficult to obtain a frame in which the number is fully visible.

However, whenever the number appears in full at some point, TTI propagates this identification correctly, and therefore it is concluded that TTI handles partial occlusions better than approaches that consider only isolated frames or do not consider the propagation of identifications over time.

Positioning

This experiment evaluates the performance of player positioning on the field. For each estimated bounding box, the corresponding ground-truth box was determined by selecting the one with the highest IoU, and the Euclidean distance between the estimated point and the actual point was computed. An estimate is considered correct if this distance is less than or equal to a certain limit, named position error tolerance. Table 6 shows the percentage of correct position estimations for different tolerance values: up to 5 meters the percentage of hits increases gradually, suggesting good accuracy. As tolerance increases, the percentage reaches a plateau, reflecting estimates that present greater deviations.

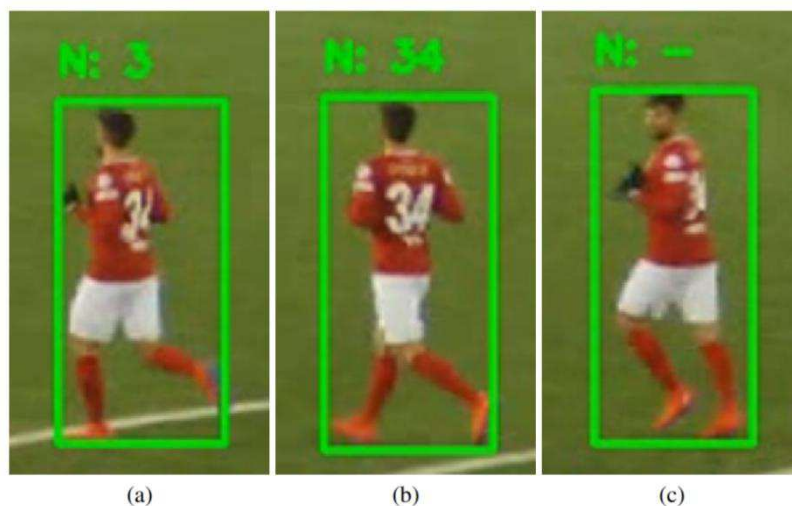


Figure 7 - TTI with Combined Digit Highlighting – spreads the identification even when the number is not visible.

A qualitative analysis using a 2D radar superimposed on the video confirmed that the estimated positions are consistent with reality, as presented in Figure 8, strengthening confidence in the proposed methodology. The biggest errors occurred in detecting the ball, especially when it was not touching the ground, generating erroneous projections due to the lack of height information.

This phenomenon of forced perspective can confuse even human observers. Positioning obtained good results, with an accuracy (hit rate) of 91.0% within a tolerance of up to 5 meters. Furthermore, the estimates generated are sufficient to infer offside situations, since it is possible to locate each player on the field with a high degree of reliability.

Table 6 - Performance of the positioning methodology at different distance error tolerances.

Position Error Tolerance (meters)	%Hits
1	66.5%
3	89.6%
5	91.0%
10	91.6%
20	92.9%
30	95.0%



Figure 8 - Example output from the positioning step – the obtained positions are resized to the 2D field representation inserted in the original image. The colors of the teams are followed in 2D field. Referees are yellow.

Comparison with SoccerNet-GSR-Based Studies

Our results can be compared with two recent studies: Somers et al. (2024) and Golovkin et al. (2025), both based on the same dataset, SoccerNet-GSR.

In the detection stage, we tested YOLOv10 and YOLOv11, obtaining the best F1-score of 83.8%. Somers et al. (2024) used YOLOv8, while Golovkin et al. (2025) employed YOLOv5. However, neither study reported detailed quantitative results for the detection stage, they only stated that these outputs were used as input for the computation of the GS-HOTA metric.

In the tracking stage, our system achieved a MOTA of 81.5%. In Somers et al. (2024), the

YOLOv8 detections were tracked in the image space using StrongSORT, but no numerical results were provided. Similarly, Golovkin et al. (2025) used DeepSORT as the tracker, also without reporting detailed quantitative results.

Regarding the GS-HOTA metric, the system presented by Somers et al. (2024) achieved approximately 22.26%, considering a maximum positional error of 5 meters. This is because the system described in the paper was still in an early development stage, focusing on complex and integrated real-time tracking and identification tasks. In contrast, the study by Golovkin et al. (2025) reached 63.81% GS-HOTA under the same 5-meter error. Although our system achieved 91% correctness in player positioning (allowing up to 5 meters of error), we acknowledge as a limitation of this study that the GS-HOTA metric was not computed.

One of the main advantages of our work is its efficiency when running on limited hardware. While Somers et al. (2024) and Golovkin et al. (2025) used NVIDIA A100 GPUs with 32 GB and 40 GB of memory, respectively, our system was developed and performed on an NVIDIA GeForce GTX 1060 with 6 GB of memory.

Conclusion

Although soccer is one of the most popular sports in the world, in-depth tactical studies remain relatively scarce due to the limited availability of data. Positional data, which are fundamental for advanced analyses, are often expensive or generated privately by teams, limiting their dissemination. This scenario makes it difficult to apply more sophisticated analytical approaches, restricting the development of evidence-based strategies. However, in recent years, initiatives that use computer vision to generate data autonomously have emerged, opening new possibilities for study. These initiatives not only enable in-depth analysis of the game but also allow the tracking of players and the ball on the field, as well as the measurement and review of game actions such as shots on goal, missed passes, and ball possession directly from match footage.

This work presents the Advanced Player Identification and Positioning System (APIPS), a comprehensive computational framework designed to estimate the spatial positions of players and the ball on the field using broadcast video footage. APIPS effectively handles challenges such as jersey number occlusions and operates under limited computational resources. The APIPS uses video frames from television broadcasts and consists of four main stages: detection, tracking, identification, and positioning. Detection locates the objects of interest (players, referees, and ball) in the images; tracking associates the detections throughout the frames, allowing the tracking of each object; identification classifies the players by team and jersey number; and positioning projects the coordinates observed in the images onto the playing field, generating a 2D matrix that represents the space of the field. This approach aims to create an efficient, low computational cost solution that can be replicated in contexts where technological resources are limited.

The results demonstrated that APIPS was able to position the players, ball, and referees in the field, achieving 91% of estimates with an error of less than 5 meters, even with modest computational resources. The detection was able to identify most objects with high accuracy and achieved an F1-score of 83.8%. The tracking showed consistency in short segments, with MOTA of 81.5%, but showed limitations in associating the same identifier to players who leave and return to the image, a problem that can be mitigated with more robust algorithms that are able to take better advantage of the visual characteristics, but which in turn require greater computational power. Team identification worked, except for goalkeepers, suggesting the inclusion of additional criteria, such as positioning on the field, to improve this classification.

The number identification was efficient but limited by the reduced visibility of the numbers in short side videos, recommending the use of longer videos or cameras positioned behind the goals to improve the frequency of frames with visible numbers. However, the TTI algorithm proved to be a useful solution to the problem of constant number occlusion, obtaining a superior result compared to the methodology that analyzes only one frame, as in Diop et al. (2022) and Alhejaily et al. (2023). The positioning, in turn, presented extremely satisfactory results, with errors restricted to situations of a moving camera or objects far from the ground, which can be corrected considering the speed of movement of the objects.

For future work, we recommend (i) computing the GS-HOTA metric to enable direct and standardized comparisons with related methods, and (ii) building a dataset with longer video sequences, which could help mitigate the identification issues observed in our experiments by providing more temporal context for tracking and re-identification. Information about the visibility of the numbers on the shirt and the height of the objects in relation to the ground would also be of great value. In addition, with the availability of greater computational resources, the refinement of each stage of the proposed methodology can be explored to improve robustness and applicability across diverse scenarios. Among the improvement possibilities, the following stands out: the use of tracking algorithms for more consistent associations through embeddings; the adaptation of the identification to include specific uniforms; and the testing of new camera and video configurations for greater precision in identifying the numbers. These advances could contribute to the system becoming a practical, accessible, and efficient tool for tactical analysis in soccer, expanding the possibilities of using positional data for both research and practical application in the sport.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Alhejaily, R., Alhejaily, R., Almdahrsh, M., Alessa, S., & Albelwi, S. (2023). Automatic Team Assignment and Jersey Number Recognition in Football Videos. *Intelligent Automation & Soft Computing*, 36(3), 2669–2684. <https://doi.org/10.32604/iasc.2023.033062>
- Barnes, C., Archer, D., Hogg, B., Bush, M., & Bradley, P. (2014). The Evolution of Physical and Technical Performance Parameters in the English Premier League. *International Journal of Sports Medicine*, 35(13), 1095–1100. <https://doi.org/10.1055/s-0034-1375695>
- Behravan, I., & Razavi, S. M. (2021). A novel machine learning method for estimating football players' value in the transfer market. *Soft Computing*, 25(3), 2499–2511. <https://doi.org/10.1007/s00500-020-05319-3>
- Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: the clear mot metrics. *J. Image Video Process.*, 2008. DOI: 10.1155/2008/246309 <https://link.springer.com/content/pdf/10.1155/2008/246309.pdf>
- Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, 3464–3468. <https://doi.org/10.1109/ICIP.2016.7533003>
- Buyrukoglu, S., & Savaş, S. (2023). Stacked-Based Ensemble Machine Learning Model for Positioning Footballer. *Arabian Journal for Science and Engineering*, 48(2), 1371–1383. <https://doi.org/10.1007/s13369-022-06857-8>

- Ciaparrone, G., Luque Sánchez, F., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381, 61–88. <https://doi.org/10.1016/j.neucom.2019.11.023>
- Cioppa, A., Giancola, S., Deliege, A., Kang, L., Zhou, X., Cheng, Z., Ghanem, B., & Van Droogenbroeck, M. (2022). SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3490–3501. <https://doi.org/10.1109/CVPRW56347.2022.00393>
- Cortez, A., Trigo, A., & Loureiro, N. (2021). Predicting Physiological Variables of Players that Make a Winning Football Team: A Machine Learning Approach. In O. Gervasi & others (Eds.), *Computational science and its applications -- iccsa 2021* (pp. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-86970-0_1
- Cortez, A., Trigo, A., & Loureiro, N. (2022). Football Match Line-Up Prediction Based on Physiological Variables: A Machine Learning Approach. *Computers*, 11(3), 40. <https://doi.org/10.3390/computers11030040>
- Cui, Y., Jiang, C., Wu, G., & Wang, L. (2024). MixFormer: End-to-End Tracking With Iterative Mixed Attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(6), 4129–4146. <https://doi.org/10.1109/TPAMI.2024.3349519>
- Cui, Y., Zeng, C., Zhao, X., Yang, Y., Wu, G., & Wang, L. (2023). SportsMOT: A Large Multi-Object Tracking Dataset in Multiple Sports Scenes. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9887–9897. <https://doi.org/10.1109/ICCV51070.2023.00910>
- Diop, C.-A., Pelloux, B., Yu, X., Yi, W.-J., & Saniie, J. (2022). Soccer Player Recognition using Artificial Intelligence and Computer Vision. *2022 IEEE International Conference on Electro Information Technology (EIT)*, 477–481. <https://doi.org/10.1109/eIT53891.2022.9813788>
- Frevel, N., Beiderbeck, D., & Schmidt, S. L. (2022). The impact of technology on sports – A prospective study. *Technological Forecasting and Social Change*, 182, 121838. <https://doi.org/10.1016/j.techfore.2022.121838>
- Giancola, S., Amine, M., Dghaily, T., & Ghanem, B. (2018). SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1792–179210. <https://doi.org/10.1109/CVPRW.2018.00223>
- Golovkin, V., Nemtsev, N., Shandyba, V., Udin, O., Kasatkin, N., Kononov, P., Afanasiev, A., Ulasen, S., & Boiarov, A. (2025). From Broadcast to Minimap: Achieving State-of-the-Art Soccernet Game State Reconstruction. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 6010–6028. <https://doi.org/10.1109/CVPRW67362.2025.00600>
- Gudmundsson, J. & Horton, M. (2017). Spatio-Temporal Analysis of Team Sports. *ACM Comput. Surv.* 50, 2, Article 22. <https://doi.org/10.1145/3054132>
- JaiedAI. (2020). *EasyOCR: Ready-to-use OCR* [Computer software]. GitHub. Accessed January 29, 2026, from <https://github.com/JaiedAI/EasyOCR>
- Jamil, M., Phatak, A., Mehta, S., Beato, M., Memmert, D., & Connor, M. (2021). Using multiple machine learning algorithms to classify elite and sub-elite goalkeepers in professional men’s football. *Scientific Reports*, 11. <https://doi.org/10.1038/s41598-021-01187-5>
- Jocher, G., & Qiu, J. (2024). *Ultralytics YOLO11* (Version 11.0.0). <https://github.com/ultralytics/ultralytics>
- King, D. E. (2009). Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.*, 10, 1755–1758. <https://dl.acm.org/doi/10.5555/1577069.1755843>

- Krishnan, K. (2013). *Data Warehousing in the Age of Big Data*. Elsevier. <https://doi.org/10.1016/C2012-0-02737-8>
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Lolli, L., Bauer, P., Irving, C., Bonanno, D., Höner, O., Gregson, W., & Di Salvo, V. (2025). Data analytics in the football industry: a survey investigating operational frameworks and practices in professional clubs and national federations from around the world. *Science and Medicine in Football*, 9(2), 189–198. <https://doi.org/10.1080/24733938.2024.2341837>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. <https://www.cs.cmu.edu/~bhiksha/courses/mlsp.fall2010/class14/macqueen.pdf>
- Manescu, D. C. (2025). Big Data Analytics Framework for Decision-Making in Sports Performance Optimization. *Data*, 10(7), 116. <https://doi.org/10.3390/data10070116>
- Majeed, F., Nazir, M., Swart, K., Agus, M., & Schneider, J. (2025). Real-time analysis of soccer ball–player interactions using graph convolutional networks for enhanced game insights. *Scientific Reports*, 15(1), 21859. <https://doi.org/10.1038/s41598-025-05462-7>
- Manish., S., Bhagat, V., & Pramila, R. (2021). Prediction of Football Players Performance using Machine Learning and Deep Learning Algorithms. *2021 2nd International Conference for Emerging Technology (INCET)*, 1–5. <https://doi.org/10.1109/INCET51464.2021.9456424>
- Memmert, D., & Rein, R. (2018). Match analysis, Big Data and tactics: current trends in elite soccer. *Deutsche Zeitschrift Für Sportmedizin*, 2018(03), 65–72. <https://doi.org/10.5960/dzsm.2018.322>
- Morra, L., Manigrasso, F., Canto, G., Gianfrate, C., Guarino, E., & Lamberti, F. (2020). Slicing and Dicing Soccer: Automatic Detection of Complex Events from Spatio-Temporal Data. In A. Campilho, F. Karray, & Z. Wang (Eds.), *Image Analysis and Recognition. ICIAR 2020. Lecture Notes in Computer Science*, vol 12131 (pp. 107–121). Springer. https://doi.org/10.1007/978-3-030-50347-5_11
- Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, 6(1), 236. <https://doi.org/10.1038/s41597-019-0247-7>
- Perl, J., Grunz, A., & Memmert, D. (2013). Tactics analysis in soccer: an advanced approach. *International Journal of Computer Science in Sport*, 12, 33–44.
- Rahman, M. A. (2020). A deep learning framework for football match prediction. *SN Applied Sciences*, 2(2), 165. <https://doi.org/10.1007/s42452-019-1821-5>
- Rampinini, E., Alberti, G., Fiorenza, M., Riggio, M., Sassi, R., Oliveira Borges, T., & Coutts, A. (2015). Accuracy of gps devices for measuring high-intensity running in field-based team sports. *International Journal of Sports Medicine*, 36. <https://doi.org/10.1055/s-0034-1385866>
- Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. CoRR, abs/1804.02767. <http://arxiv.org/abs/1804.02767>
- Sarmento, H., Clemente, F. M., Araújo, D., Keith, D., McRobert A., & Figueiredo, A. (2018). What Performance Analysts Need to Know About Research Trends in Association Football (2012–2016): A Systematic Review. *Sports Med* 48, 799–836. <https://doi.org/10.1007/s40279-017-0836-6>
- Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(11), 2298–2304. <https://doi.org/10.1109/TPAMI.2016.2646371>

- Scott, A., Uchida, I., Ding, N., Umemoto, R., Bunker, R., Kobayashi, R., Koyama, T., Onishi, M., Kameda, Y., & Fujii, K. (2024). TeamTrack: A Dataset for Multi-Sport Multi-Object Tracking in Full-pitch Videos. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3357–3366. <https://doi.org/10.1109/CVPRW63382.2024.00340>
- Somers, V., Joos, V., Cioppa, A., Giancola, S., Ghasemzadeh, S. A., Magera, F., Standaert, B., Mansourian, A. M., Zhou, X., Kasaei, S., Ghanem, B., Alahi, A., Van Droogenbroeck, M., & De Vleeschouwer, C. (2024). SoccerNet Game State Reconstruction: End-to-End Athlete Tracking and Identification on a Minimap. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3293–3305. <https://doi.org/10.1109/CVPRW63382.2024.00334>
- Theiner, J., & Ewerth, R. (2023). TVCalib: Camera Calibration for Sports Field Registration in Soccer. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1166–1175. <https://doi.org/10.1109/WACV56688.2023.00122>
- Theiner, J., Gritz, W., Muller-Budack, E., Rein, R., Memmert, D., & Ewerth, R. (2022). Extraction of Positional Player Data from Broadcast Soccer Videos. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1463–1473. <https://doi.org/10.1109/WACV51458.2022.00153>
- Vashist, M., Bahl, V., Goel, A., & Sengar, N. (2021). Full time result prediction using ensemble techniques. *Asian Journal of Convergence in Technology*, 7, 38–42. <https://doi.org/10.33130/AJCT.2021v07i03.006>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-time end-to-end object detection. *Proceedings of the 38th International Conference on Neural Information Processing Systems (NeurIPS '24)*, Vol. 37. Curran Associates Inc., Red Hook, NY, USA, Article 3429, 107984–108011. <https://dl.acm.org/doi/10.5555/3737916.3741345>
- Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 3645–3649. <https://doi.org/10.1109/ICIP.2017.8296962>
- Wright, C., Atkins, S., & Jones, B. (2012) An analysis of elite coaches' engagement with performance analysis services (match, notational analysis and technique analysis). *International Journal of Performance Analysis in Sport*. 436–451. <https://doi.org/10.1080/24748668.2012.11868609>
- Yang, C., Yang, M., Li, H., Jiang, L., Suo, X., Mao, L., Meng, W., & Li, Z. (2025). A survey on soccer player detection and tracking with videos. *The Visual Computer*, 41(2), 815–829. <https://doi.org/10.1007/s00371-024-03367-6>
- Yang, W. (2021). Predict soccer match outcome based on player performance. *Frontiers in Sport Research*, 3(3), 74–78. <https://doi.org/10.25236/FSR.2021.030314>
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object Tracking by Associating Every Detection Box. In *Computer Vision -- ECCV 2022. Lecture Notes in Computer Science*, vol 13682 (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-20047-2_1