Ensino que te conecta

BACHELOR IN
SOFTWARE ENGINEERING

DETECTION AND ANALYSIS OF TACTICAL FORMATIONS IN SOCCER USING COMPUTER VISION

VITOR SOUZA OLIVEIRA

Brasília - DF, 2025

VITOR SOUZA OLIVEIRA

DETECTION AND ANALYSIS OF TACTICAL FORMATIONS IN SOCCER USING COMPUTER VISION

Undergraduate Thesis presented as a partial requirement for the degree of Bachelor in Software Engineering, at the Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP).

Advisor
Prof. Dr. Rafael Lemos Paes

Brasília - DF, 2025

Código de catalogação na publicação – CIP

O48d Oliveira, Vitor Souza

Detection and Analysis of Tactical Formations in Soccer Using Computer Vision / Vitor Souza Oliveira — Brasília: Instituto Brasileiro Ensino, Desenvolvimento e Pesquisa, 2025.

54 f. : il.

Orientador: Prof. Dr. Rafael Lemos Paes

Monografia (Bacharelado em Engenharia de Software) - Instituto Brasileiro Ensino, Desenvolvimento e Pesquisa – IDP, 2025.

1. Computer vision. 2. Object detection. 3. Soccer tactical analysis. I. Título

CDD 006.3

Elaborada por Biblioteca Ministro Moreira Alves

VITOR SOUZA OLIVEIRA

DETECTION AND ANALYSIS OF TACTICAL FORMATIONS IN SOCCER USING COMPUTER VISION

Undergraduate Thesis presented as a partial requirement for the degree of Bachelor in Software Engineering, at the Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP).

Approved on 12/04/2025

Examining Committee

Documento assinado digitalmente
RAFAEL LEMOS PAES
Data: 19/12/2025 21:34:56-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Rafael Lemos Paes - Advisor

Documento assinado digitalmente
LUCAS MAURICIO CASTRO E MARTINS
Data: 20/12/2025 13:50:11-0300
Verifique em https://validar.iti.gov.br

Prof. Me. Lucas Maurício Castro e Martins - Internal Examiner

ASSINADO DIGITALMENTE
FABRICIO FERNANDES SANTANA
A conformidade com a assinatura pode ser verificada em:
http://serpro.gov.br/assinador-digital

Prof. Fabrício Fernandes Santana - Internal Examiner

DEDICATION

To my parents, José Gilson and Roseli de Souza, to my sister, Luiza de Souza, and to my girlfriend, Letícia Macedo.
And to God, for making all of this possible.

ACKNOWLEDGMENTS

First, I thank God for granting me health, wisdom, and the strength necessary to overcome the challenges of this journey and make it this far.

To my parents, José Gilson de Oliveira and Roseli de Souza de Oliveira, I offer my eternal gratitude. Thank you for providing all the emotional and material support throughout my course, for every sacrifice made for my education, and for always believing in my future.

A special thanks to my sister, Luiza de Souza de Oliveira. Your help was the decisive step that allowed me to enter this institution and begin this dream. I am grateful to you for opening the doors to my professional development.

To my girlfriend, Letícia Macedo, thank you for being my daily partner in the pursuit of knowledge. Your constant support drives me to evolve and become a better person. I am grateful for your patience and for walking beside me at every stage of this achievement.

To my advisor, Prof. Dr. Rafael Lemos Paes, I am grateful for the technical guidance in this work, which was fundamental to my growth as a Software Engineer.

Finally, I would like to thank the professors and colleagues at IDP who contributed to my technical and human development by providing a collaborative and enriching learning environment.

RESUMO

A análise tática no futebol enfrenta um impasse estrutural: enquanto ferramentas tecnológicas avançadas apresentam custos elevados, processos manuais continuam predominantes devido à ausência de soluções automatizadas acessíveis, sobrecarregando analistas, especialmente em clubes com menor investimento. Este trabalho busca mitigar essa lacuna ao propor e validar uma metodologia de baixo custo para a identificação automática de formações táticas a partir de vídeos. A abordagem emprega visão computacional para processar imagens provenientes de câmeras táticas ou transmissões televisivas, integrando o detector de objetos YOLOv8 ao algoritmo de rastreamento BoT-SORT, cuja capacidade de compensação da movimentação da câmera é fundamental para a robustez do sistema. A classificação das equipes é realizada por meio de segmentação cromática no espaço de cor CIELAB com K-Means, enquanto a inferência tática é obtida por um modelo estatístico dinâmico que mapeia as linhas de defesa, meio-campo e ataque. A validação, conduzida por meio da comparação entre as formações inferidas e as escalações oficiais das partidas, demonstrou a eficácia da metodologia proposta. Os resultados evidenciam um caminho concreto para democratizar a análise de desempenho, reduzindo a dependência de sistemas proprietários e oferecendo maior autonomia às comissões técnicas.

Palavras-chave: Visão Computacional, Detecção de Objetos, Rastreamento de Objetos, Análise Tática no Futebol, Futebol.

ABSTRACT

Tactical analysis in soccer faces a structural dilemma: advanced technological tools entail high operational costs, while manual processes still dominate due to the lack of accessible automation, placing a significant burden on analysts, particularly in lower-budget clubs. This work addresses this gap by proposing and validating a low-cost methodology for the automated identification of tactical formations from video footage. The method employs computer vision to process images from tactical cameras or broadcast feeds, combining the YOLOv8 object detector with the BoT-SORT tracking algorithm, selected for its robustness and ability to compensate for camera motion. Team classification is performed through color-based segmentation in the CIELAB color space using K-Means, while tactical inference is achieved through a dynamic statistical model that estimates defensive, midfield, and attacking lines. Validation was conducted by comparing the automatically inferred formations with official match lineups, demonstrating the effectiveness of the proposed approach. The results indicate a viable path toward democratizing performance analysis by reducing the dependence on proprietary systems and streamlining the workflow of coaching staffs.

Keywords: Computer Vision, Object Detection, Object Tracking, Soccer Tactical Analysis, Soccer.

CONTENTS

1

1 INTRODUCTION

1.1 CONTEXTUALIZATION

For decades, sports performance analysis was a field dominated by the intuition of coaches and the qualitative observation of scouts [1]. However, this paradigm experienced a clear turning point, popularized by the book *Moneyball* [2]. The book chronicled how statistical analysis was used to subvert traditional baseball logic, demonstrating that evidence-based decisions could outperform intuition. This philosophy rapidly inspired a data rush across various other sports; however, its application to soccer proved to be a unique challenge due to the game's fluid and tactically complex nature [3].

To understand the depth of this transformation beyond mere statistics, the work presented in [4] provides excellent context. The authors define the current landscape as the "digitization of sports," a phenomenon that not only optimizes performance but also fundamentally transforms club structures. According to them, this shift directly impacts the technological sector—through the proliferation of sensors and cameras—and the organizational sector, marked by the entry of new actors such as software companies and the necessity for new management models.

In this context of sports redefined by technology, Computer Vision stands out as a high-potential tool. Its ability to analyze the game collectively by tracking player positioning enables the quantification of tactical analysis, transforming subjective perceptions of team organization into objective, measurable metrics. The application of similar technologies in elite tournaments, such as the FIFA World Cup, as evidenced in [4], demonstrates the relevance of this topic. However, the same research highlights that this is still a nascent field, presenting a clear opportunity for investigations such as the one proposed in this work.

1.2 THE SPECIFIC PROBLEM: THE TECHNOLOGICAL ACCESSIBILITY GAP

Despite the growing importance of data analysis, access to technological tools remains limited for a significant portion of soccer professionals. Clubs and analysts operating outside the financial elite face a significant dilemma: investing in high-cost, cutting-edge software or resorting to manual processes, which are inherently slower, laborious, and subjective.

This dilemma manifests concretely in the daily routine of industry professionals. In an interview conducted for this work with the performance analyst of Capital FC, a team in the Brazilian Série D, the practical challenges faced on a daily basis were revealed. The choice of this institution was due to geographic accessibility and professional recommendation, allowing for requirements gathering within a real competition scenario.

During data collection, performed via audio records with the analyst (as detailed in Appendix B), it was reported that the department consists of only one professional, creating operational difficulties given the scarcity of information in Série D and the regional championship. The analyst highlighted that, although the club uses GPS technology solely for physiological data, there is no structured data collection process for tactical performance aimed at future applications, such as Machine Learning.

The current workflow relies on *Nacsport* software for video organization, but the analysis itself requires a manual tagging process. According to the interviewee, the team relies on observing three opponent matches (home, away, and after substitutions) to detect offensive, defensive, and set-piece patterns. The professional mentioned other tools in the market, such as the *Hudl* platform, but cited difficulties in implementing previous technologies due to the low reliability of third-party data available for this division.

Regarding metrics, the analyst emphasized the importance of specific collective indicators requested by the coaching staff, such as quantifying the ”times the team entered the final third” and identifying ”through which channel” this progression occurred. Currently, obtaining this data depends on human observation, consuming time that could be dedicated to strategy.

To contextualize the scale of this challenge, a survey of the main available software solutions was conducted. The comparative analysis, considering functionalities, costs, and the target audience of each tool, is consolidated in Table 1.

Table 1 – Comparison of commercial software for soccer tactical analysis

Software	Primary Function	Cost
Nacsport [5]	Tactical analysis and video coding (tagging). Interactive dashboards, presentations, and live analysis.	Ranging from US$ 165/year (Basic) to US$ 2,590/year (Elite)
Sportscode [6]	Industry standard for high-performance video coding. Tactical analysis and report generation.	Not publicly disclosed
Wyscout (Hudl) [7]	Scouting and recruitment. Global player/video database for market and opponent analysis.	Not publicly disclosed
Catapult Vision [8]	Video analysis integrated with physical performance data (GPS). Connects tactical to physical aspects.	Not publicly disclosed
StatsBomb IQ [9]	Event data analysis platform (non-video). Focus on data science and advanced statistics.	Not publicly disclosed
Metrica Sports [10]	Video analysis with tracking tools and tactical visualizations. Good cost-benefit ratio.	Ranging from € 60/month to undetermined (higher-tier plan)

Source: The author.

Table 1, therefore, not only illustrates but underscores the gap dividing the market. On one hand, elite solutions such as Hudl and StatsBomb offer immense analytical power, yet they come with undisclosed costs and contracts inaccessible to most organizations. On the other hand, entry-level solutions like Nacsport address video organization but still impose a manual workload on the analyst for tactical data extraction.

To gauge the financial barrier imposed by these solutions, it is important to contextualize the costs against the budgetary reality of lower-investment clubs. Using the Nacsport software (Elite version) as a benchmark, with an annual cost of approximately US$ 2,590 (as per Table 1), it is observed that the investment for a single license represents a significant sum for clubs with limited resources.

In contrast, the fixed participation fee paid by the Brazilian Football Confederation (CBF) to Série D clubs in 2025 was set at a value equivalent to approximately US$ 83,295 for the entire first phase of the competition [11]. Considering that this resource must fund about four months of operations, including logistics, food, and the entire squad's payroll, the acquisition of a single software license would consume a disproportionate fraction of the club's guaranteed budget. Such financial commitment becomes prohibitive given basic operational priorities and establishes cost as an impediment to technological innovation in these institutions.

It is precisely within this gap of accessible tools for automated tactical analysis via Computer Vision that the need to investigate a new solution arises, which is the central proposal of this work.

1.3 THE PROPOSED SOLUTION AND JUSTIFICATION

Identifying an innovation opportunity in the sports market, this project is dedicated to developing and validating a methodology to automate the analysis of tactical formations in soccer using Computer Vision. The central proposal consists of engineering a software artifact capable of processing both tactical camera footage and videos from television broadcasts (broadcast).

The solution aims to map athlete coordinates and, through statistical modeling, infer the team's tactical lines (defense, midfield, and attack) and classify the predominant formation (e.g., 4-4-2) and its dynamic variations throughout the match.

The justification for this work lies in the democratization of access to technology. Given that current high-end tools impose prohibitive financial barriers on lower-investment clubs, the development of a solution based on efficient and accessible algorithms presents itself as a viable path to optimize the workflow of performance analysts at all levels of the sport.

1.4 RESEARCH QUESTION AND HYPOTHESIS

Considering the challenge of transforming the analyst's workflow, allowing for greater focus on strategic interpretation and reduced effort on manual data collection, this work is guided by the following central question:

Is it possible, through the integration of object detection and tracking techniques, to automate the identification of tactical formations in soccer videos with validatable accuracy and low computational cost?

To answer this question, the hypothesis is formulated that the combination of modern convolutional neural networks with motion-compensated tracking algorithms and temporal stabilization heuristics enables the development of a methodology capable of inferring a team's tactical structure. It is expected to demonstrate that such an approach is technically viable for post-match analysis applications, overcoming the cost barriers of proprietary solutions.

1.5 OBJECTIVES

1.5.1 General

To develop an automated methodology, based on Computer Vision techniques, for the mapping and identification of tactical formations in soccer videos, aiming to optimize data extraction and support performance analysts' decision-making.

1.5.2 Specifics

- **(SO1)** Investigate the accessibility limitations of commercial performance analysis tools regarding the budgetary reality of lower-investment clubs;

- **(SO2)** Prepare the database for model training by utilizing and processing public soccer image datasets;
- **(SO3)** Implement a Computer Vision pipeline capable of performing player detection, temporal tracking, and tactical zone inference;
- **(SO4)** Measure the effectiveness of the system's tactical inference by comparing the automatically generated results against official match lineups.

1.6 DOCUMENT STRUCTURE

This study is organized into five chapters, structured to guide the reader from theoretical background to practical validation. Chapter 1 introduces the research problem and objectives, followed by Chapter 2, which provides the theoretical foundation for performance analysis concepts and reviews the state of the art in detection and tracking algorithms. Chapter 3 details the solution engineering, describing the pipeline architecture and the mathematical model developed for tactical inference. Subsequently, Chapter 4 presents the experimental protocol and discusses the results obtained during the validation of the proposed method. Finally, Chapter 5 synthesizes the study's contributions, addresses its limitations, and outlines directions for future work.

2

2 LITERATURE REVIEW

2.1 THEORETICAL BACKGROUND

2.1.1 Performance Analysis and Digital Transformation in Soccer

As a field of knowledge within Sports Science, Performance Analysis consists of the process of systematic observation and recording of events, seeking to identify patterns and factors that influence performance with the ultimate goal of optimizing decision-making to enhance sporting success [1]. Although early studies of this genre date back to the early 20th century, it was the combination of professionalization and technological advancement, starting in the 1980s, that truly consolidated the field in elite soccer [12].

The professional at the center of this entire operation is the Performance Analyst, whose role extends far beyond merely filming or compiling statistics. The study in [12], interviewing ten analysts from Brazilian Série A clubs, reveals the complexity of this professional's routine, which ranges from footage acquisition to clip editing and conducting individual and collective analyses. This process of providing video feedback, especially to young athletes, is internationally recognized as a powerful pedagogical tool, yet one that requires great care to avoid negatively impacting player confidence [13].

Despite the growing importance of the field, the analyst's routine is marked by concrete challenges. Research in [12] points to work overload and inadequate infrastructure as major obstacles in the Brazilian scenario. This reality forces many to use personal equipment and rely on manual video tagging processes, a task described as time-consuming and laborious. Such challenges are consistent with findings in international literature, which also identify coaches' lack of time and the difficulty in building a positive learning environment as barriers to the effectiveness of performance analysis [13].

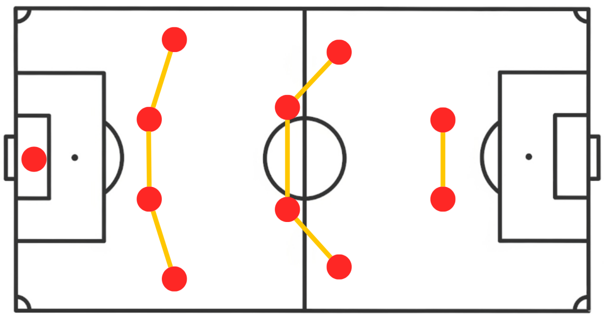
Historically, analysis tools have evolved from manual notations to the support of computerized systems that automate information recording [12]. However, a persistent bottleneck exists: interpreting data to identify complex tactical patterns, such as team organization and movements, remains a task that relies largely on manual intervention and the analyst's trained eye. This dependency on the human factor, combined with the high cost of cutting-edge technologies, highlights the need for solutions that advance toward automating tactical analysis itself, making it faster, objective, and accessible.

The current landscape of sports data analysis fits into a much larger movement, defined by academia as the “digitization of sports.” As contextualized in [4], this concept describes a profound transformation in which technology not only optimizes performance but reconfigures the very foundations of the sport, from its organization to how it is consumed. One of the turning points that catalyzed this data rush was the Moneyball philosophy. Popularized by the work [2], its premise was radical: using statistical analysis to build competitive teams more intelligently, relying on evidence rather than traditional intuition. Although this thinking sparked a data race in the sporting world, its application to soccer proved to be a unique challenge. The fluid and tactically interdependent nature of the game makes it difficult to isolate indicators that lead directly to victory, a point already analyzed in detail in [3].

2.1.2 Fundamental Tactical Concepts in Soccer

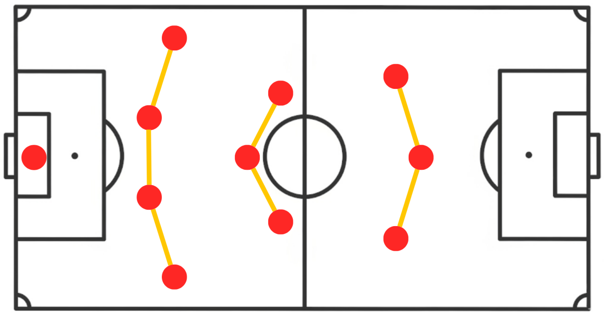
Modern soccer tactical analysis investigates a team’s collective organization, fundamentally represented by its playing system, also known as tactical formation [1]. This formation is a numerical representation, such as 4-4-2 or 4-3-3, describing the distribution of players across defensive, midfield, and offensive lines, defining their initial spatial responsibilities on the pitch. From a computational perspective, the formation can be understood as the team structure, a positioning pattern that emerges from players’ spatiotemporal data throughout the match [14]. This structure defines how the team organizes itself to maintain possession, create scoring opportunities, and impede opponent advancement.

Figure 1 – Representation of a 4-4-2 tactical system



Source: The author.

Figure 2 – Representation of a 4-3-3 tactical system



Source: The author.

However, it is important to distinguish between the *nominal formation* and the *dynamic formation*. The nominal formation is the static playing system, planned by the coaching staff and announced prior to the match. Nevertheless, due to the game's fluid nature, this structure rarely remains fixed. The dynamic formation, conversely, corresponds to the athletes' actual average positioning, which can be extracted and analyzed using high-frequency positional data [15]. The ability to discover and analyze these dynamic structures, representing the team's actual tactical behavior, is a primary objective of modern performance analysis approaches, going beyond the simplicity of the nominal formation [14].

The variation between the nominal and dynamic formations is directly influenced by the different phases of play, which are the four distinct moments characterizing a match: offensive organization (when the team possesses the ball and builds an attack), defensive transition (the immediate moment following possession loss), defensive organization (when the team positions itself to defend without the ball), and offensive transition (the immediate moment following possession recovery) [1]. A team's spatial organization changes drastically during each of these phases; for instance, a team may defend in a compact 4-4-2 formation but shift into a 4-2-4 during the attacking phase to maximize offensive presence. Therefore, tactical analysis is not limited to identifying a single formation but involves understanding how the team structure dynamically adapts to each phase of the game.

2.1.3 Computer Vision Technologies

Extracting tactical knowledge from raw video is a complex computational challenge requiring the orchestration of state-of-the-art algorithms. The process can be divided into two fundamental and sequential stages: first, detecting all players in each video frame, and second, tracking these players over time to maintain their identity and construct their trajectories. In this new scenario, Machine Learning assumes a central role. Its ability to process complex information to identify patterns invisible to human analysis opens a new horizon of possibilities.

2.1.4 Neural Networks, Deep Learning, and Convolutional Specialization

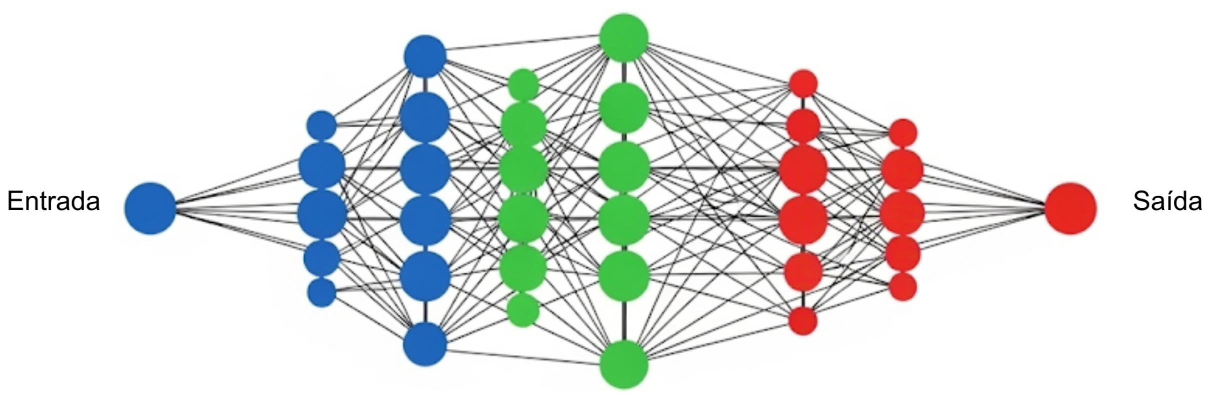
The core technologies of this work, YOLO and DeepSORT, are applications of an Artificial Intelligence subfield known as Deep Learning. Deep Learning utilizes Artificial Neural Network (ANN) architectures with multiple layers to learn data representations hierarchically and automatically.

A neural network, in its most basic form, is a computational model inspired by the human brain structure, composed of interconnected nodes, or neurons, organized in layers (input, hidden, and output). Each neuron receives input signals, processes them, and passes an output signal to the neurons of the subsequent layer. Mathematically, the output y of a neuron can be represented as:

$$y = f(\sum_{i=1}^{n}(w_i \cdot x_i) + b)$$

Where x_i are the inputs, w_i are the synaptic weights (parameters the network learns during training), b is a bias, and f is a non-linear activation function. A network becomes ”deep” when it possesses multiple hidden layers, enabling the learning of increasingly complex and abstract patterns.

Figure 3 – Conceptual representation of a Neural Network architecture.



Source: The author.

For computer vision tasks, such as detecting players on a soccer pitch, a specialized class of deep neural network has proven orders of magnitude more effective: the Convolutional Neural Network (CNN). As employed in architectures such as YOLO [16], a CNN is designed to process data with a grid topology, such as an image. Its main characteristic is the convolution operation.

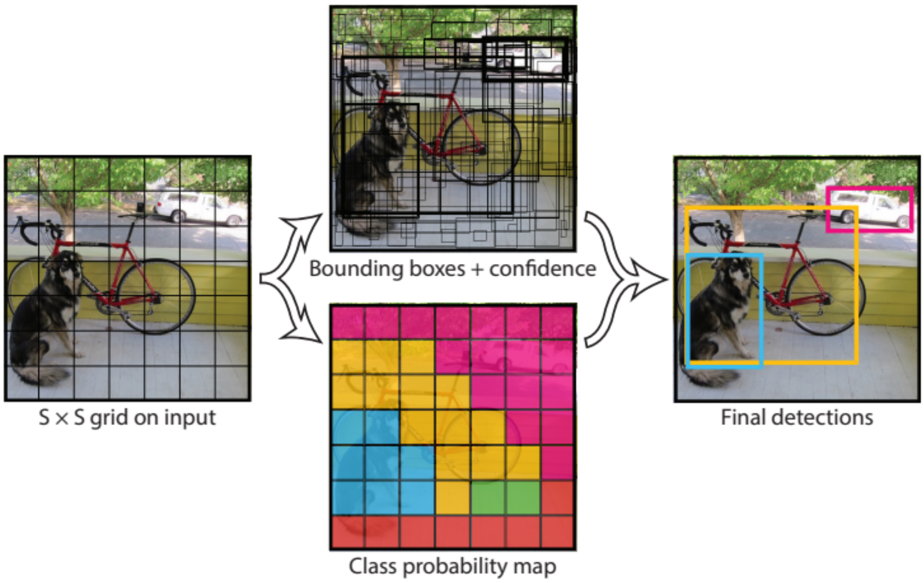
Unlike a traditional neural network, where each neuron connects to all neurons in the previous layer, a CNN utilizes filters (or kernels) that slide over the input image. Each filter is a small feature detector, specialized in identifying simple patterns such as edges, colors, or textures. As filters traverse the image, they generate ”feature maps” indicating where these patterns were found. Subsequent convolutional layers combine these maps to detect more complex patterns, such as the outline of a person or the shape of a ball. This process, combined with pooling layers that reduce spatial dimensionality, allows the CNN to learn a hierarchical representation robust to scale and translation variations, making it the ideal architecture for object detection tasks.

2.1.5 Object Detection with the YOLO Family

Object detection, the task of locating and classifying object instances in an image, is the first pillar of the proposed solution. Modern approaches utilize Convolutional Neural Networks (CNNs), and among them, the YOLO (You Only Look Once) model family stands out as the state of the art for real-time detection. Originally proposed in the seminal work [16], YOLO revolutionized the field by treating detection as a single regression problem. Unlike two-stage models (such as the R-CNN family), it analyzes the image only once to predict bounding boxes and class probabilities, resulting in significantly superior processing speed.

Over the years, YOLO has evolved into several versions. YOLOv3 solidified the architecture as a market standard, offering an excellent balance between speed and accuracy. More recent versions such as YOLOv5 and YOLOv8 (developed by the company Ultralytics) focused on improving not only performance but also the ease of implementation and training. For the development of a prototype such as the one proposed in this undergraduate thesis, the choice of a modern version like YOLOv8 is strategic, as it offers very high performance, vast documentation, and, crucially, the availability of pre-trained models, which drastically accelerates the development cycle and ensures the project’s viability.

Figure 4 – YOLO Detection Model with Bounding Boxes



Source: Adapted from [16].

The operation of YOLO relies on superimposing an S$\times$S grid over the input image. If the center of an object falls within a specific grid cell, that cell becomes responsible for detecting that object. Each grid cell predicts B bounding boxes and a confidence score for each one. This score reflects the certainty that the box contains an object and the accuracy with which it delimits it. Additionally, each cell predicts a set of class probabilities, which, combined with the bounding boxes, allow for the classification of the detected objects. To evaluate the accuracy of a predicted box, the Intersection over Union (IoU) metric is used, which calculates the ratio between the area of intersection of the predicted box with the ground truth box and the total area of their union, formally expressed as:

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$

Over the years, YOLO has evolved into several versions. YOLOv3 solidified the architecture as a market standard, offering an excellent balance between speed and accuracy. More recent versions such as YOLOv5 and YOLOv8 (developed by the company Ultralytics) focused on improving not only performance but also the ease of implementation and training. For the development of a prototype such as the one proposed in this undergraduate thesis, the choice of a modern version like YOLOv8 [17] is strategic, as it offers very high performance, vast documentation, and, crucially, the availability of pre-trained models, which drastically accelerates the development cycle and ensures the project's viability.

2.1.6 Transfer Learning

Training deep Convolutional Neural Networks from scratch demands large volumes of annotated data and high computational power, resources not always available in specific applications. To mitigate this problem, the Transfer Learning technique is employed.

According to [18], transfer learning consists of a system's ability to recognize and apply knowledge learned in a source domain (with abundant data) to improve learning in a target domain (where data may be scarce). In the context of computer vision, this is possible because the early layers of a CNN learn generic features (such as edges, textures, and simple shapes) that are common to almost all visual images.

Thus, it is possible to utilize the weights of a network pre-trained on a massive, generic dataset, such as MS COCO (Microsoft Common Objects in Context), and perform only the fine-tuning of the final layers to detect specific objects, such as soccer players. This approach drastically accelerates training convergence and improves model generalization, preventing overfitting on smaller datasets.

2.1.7 Multi-Object Tracking with DeepSORT

Following player detection in a frame, the challenge arises of ensuring that a player detected in frame t is recognized as the same individual in frame $t+1$. This task is known as Multi-Object Tracking (MOT). One of the most well-known and effective algorithms for this task is DeepSORT, proposed in the work [19].

The SORT component of the algorithm utilizes a Kalman Filter, a classic recursive estimation algorithm [20], to predict the future position of each player based on their current movement. The Kalman Filter is optimized to estimate a system's state from a series of noisy measurements, such as the bounding box detections provided by YOLO. It operates in a two-stage recursive cycle: prediction, where the player's future state (position and velocity) is estimated based on the current state, and update, where the prediction is corrected using the new measurement (the detection in the current frame), resulting in a more accurate and smoothed trajectory estimate. The system state can be modeled by the following simplified equations:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_k \quad \text{(State transition equation)}$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \quad \text{(Measurement equation)}$$

Where $\mathbf{x}$ represents the player's state (containing position and velocity coordinates), $\mathbf{F}$ is the transition matrix modeling the physics of motion, and $\mathbf{z}$ is the observed measurement (the YOLO bounding box). The terms $\mathbf{w}$ and $\mathbf{v}$ represent the process and measurement noise, respectively.

Based on this motion prediction, the algorithm performs data association, efficiently connecting YOLO detections in the new frame with the predicted trajectories. The major innovation, the ”Deep” in the name, is the use of a pre-trained deep neural network to extract a feature vector (embedding) representing each player’s unique appearance. When a player is occluded and reappears, the system does not rely solely on motion but compares the new player’s visual signature with those who disappeared, enabling robust re-identification and the maintenance of the same ID for the player.

Despite its effectiveness, DeepSORT possesses limitations, particularly in scenarios with heavy occlusions, players with highly similar appearances (identical uniforms), and videos with significant camera movement, where Kalman Filter predictions become less reliable.

2.1.8 The State of the Art in Tracking: BoT-SORT

To overcome the limitations observed in trackers based purely on linear Kalman Filters, such as DeepSORT, recent literature has advanced toward architectures more robust to dynamic scenarios. In this context, BoT-SORT (Bolstered Tracking-by-Detection), proposed in [21], stands out.

The main contribution of this approach is the integration of Camera Motion Compensation (CMC) into the tracking process. In sports broadcasts, constant camera movement (pan, tilt, and zoom) introduces a global displacement in the image that confounds player motion predictions. BoT-SORT utilizes image feature alignment to calculate the homography matrix between adjacent frames, correcting bounding box coordinates and allowing the Kalman Filter to operate on the object’s actual motion rather than the apparent motion induced by camera movement.

Beyond CMC, the algorithm enhances data association by combining IoU (Intersection over Union) distance and cosine distance (visual re-identification) with stricter motion metrics, resulting in a significant reduction of identity switches (ID switches) and trajectory fragmentation, essential characteristics for the continuous tactical analysis proposed in this work.

2.2 RELATED WORK

2.2.1 Artificial Intelligence in Soccer

The application of Artificial Intelligence, and more specifically Machine Learning, in soccer is a consolidated and rapidly expanding research field. Recent systematic reviews, such as [22], help organize this landscape, showing that studies cluster into three major areas: injury prediction, talent identification, and, most relevant to this work, performance analysis.

Within performance analysis, Computer Vision has emerged as one of the most powerful tools. The review [23] maps the vast array of applied technologies, ranging from event detection and player action classification [24] to tactical behavior analysis, validating the area's relevance and potential. To analyze the studies connecting more directly to this work, the following discussion narrows the focus to the challenges and applications of computer vision in extracting tactical metrics.

2.2.2 Approaches for Tactical Analysis via Computer Vision

The foundation for any automated tactical analysis from video is the ability to detect and, crucially, track the identity of each player over time. This challenge, known as Multi-Object Tracking (MOT), is notoriously complex in the soccer context. A fundamental work highlighting this complexity is [25], which introduced SoccerNet-Tracking, a large-scale benchmark for MOT in soccer videos. The study's conclusion was unequivocal: even the best algorithms of the time presented significant difficulties in scenarios of occlusion and rapid movement, validating that the tracking stage is, in itself, a non-trivial engineering problem.

Once player trajectories are extracted, the literature divides into various applications. In [26], for example, the authors proposed an automated system to generate tactical performance statistics focusing on the individual. Utilizing a CNN pipeline, the system extracts metrics such as ball possession and pass counts, demonstrating the feasibility of automating data collection. In contrast, the seminal work [14] focuses on collective structure analysis, proposing an unsupervised learning method to discover tactical formations from high-precision positional data, establishing a gold standard for tactical analysis, albeit dependent on high-cost technology.

Looking at the research frontier, the review [27] analyzes the application of even more advanced Deep Learning architectures, such as Transformers, to capture complex spatiotemporal interactions between players. This strategically positions the present work: focused on a fundamental problem (formation identification), utilizing robust and accessible technologies (YOLO and DeepSORT), while cutting-edge research advances toward predictive models of greater complexity.

2.2.3 Comparative Analysis and Gap Synthesis

The synthesis of related work, consolidated in Table 2, reveals a division in the state of the art. On one hand, academic literature validates the technical complexity of multi-object tracking in occlusion scenarios, frequently proposing high-complexity predictive models requiring robust hardware. On the other, existing commercial solutions focus primarily on individual statistics or rely on proprietary infrastructure to generate collective tactical data.

Table 2 – Comparative Analysis of Related Work in Tactical Analysis via Computer Vision

Work	Main Objective	Methodology	Data Source	Contribution / Focus
Cioppa *et al.* [25]	Create a benchmark for player tracking (MOT).	Large-scale annotation of match videos for model evaluation.	High-quality broadcast videos.	Validates tracking as a fundamental technical challenge and provides data to the community.
Bialkowski *et al.* [14]	Automatically discover tactical formations.	Unsupervised learning (clustering).	High-precision positional data (multi-camera systems).	Defines the gold standard for tactical analysis, but with high-cost and inaccessible technology.
Theagarajan & Bhanu [26]	Automate performance statistics generation.	CNN pipeline for detection, tracking, and ball possession.	Broadcast videos.	Focus on individual performance (passes, possession), not the team's collective structure.
Huang *et al.* [27]	Review the state of the art of advanced DL models.	Analysis of works utilizing architectures such as Transformers.	Positional and video data.	Points to the future of predictive analysis, highlighting the complexity of frontier models.
This Work	**Identify and analyze tactical formations in an automated and accessible manner.**	**Computer Vision (YOLOv8 + BoT-SORT).**	**Tactical camera videos (low cost).**	**Focus on collective structure, seeking tactical analysis depth with accessible and open-source methods.**

Source: The author.

It becomes evident, therefore, that there is an absence of a solution addressing a central problem for the vast majority of clubs: the automated identification of the team's tactical structure without relying on wearable sensors or dedicated camera systems. The identified gap is not merely functional, but one of accessibility. The market lacks a tool that converts raw broadcast footage into actionable tactical intelligence, serving as a foundation for more complex analyses.

Given this scenario, the strategic positioning of this work distinguishes itself not by seeking absolute accuracy in laboratory conditions, but by the feasibility of implementation in resource-constrained scenarios. While elite tools (such as StatsBomb and Hudl) impose financial and operational barriers, the proposed solution is grounded in technological democratization. The competitive differential lies in reducing operational costs and eliminating the need for dedicated capture infrastructure, aligning directly with the operational constraints of clubs operating outside the financial elite.

3

3 METHODOLOGY

The methodological framework presented herein organizes the stages of conception, development, and validation of the proposed computational system. Priority is placed on the technical description of the video processing pipeline, justifying the selection of algorithms responsible for converting unstructured data into actionable tactical information.

Regarding its nature, the research is classified as applied and experimental, grounded in the engineering of a computational artifact for the automation of tactical analysis, in accordance with the taxonomy proposed by [28]. The adopted approach is characterized as mixed, combining qualitative assessments of the tracking algorithms' visual robustness with the quantitative measurement of the convergence rate between the inferred tactical formation and the official nominal strategy. The experimental protocol relied on the controlled manipulation of hyperparameters, including confidence thresholds and voting windows, aiming to optimize the fidelity of the tactical representation.

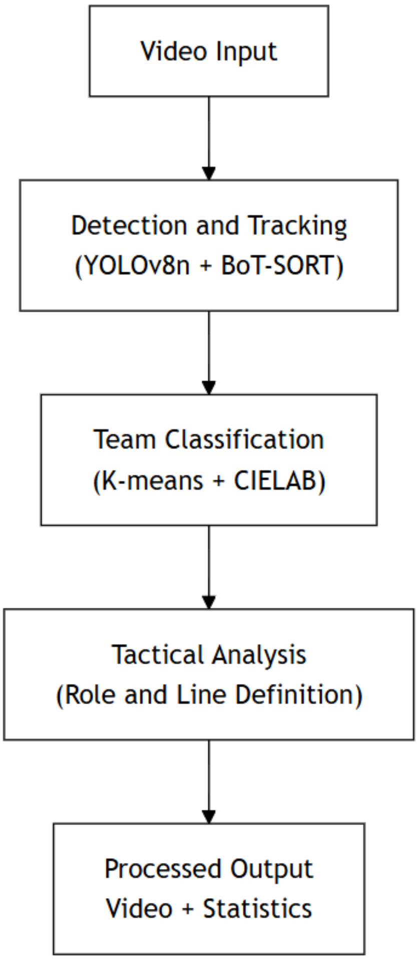
To ensure scientific reproducibility and logical clarity, the chapter is organized following the application's data flow: it begins with the research classification and high-level architecture, delves into detection, team segmentation, and temporal tracking techniques, and culminates in the formalization of the statistical model developed for the dynamic inference of tactical formations.

3.1 PROCESSING PIPELINE ARCHITECTURE

The system architecture follows a modular computer vision design organized into a sequential pipeline. This separation aims to isolate the visual feature extraction stage from the tactical inference module, ensuring decoupling between perception (vision) and business logic (tactics). Such architecture allows for the independent adjustment of hyperparameters for each phase, as well as facilitates the interchangeability of predictive models without requiring refactoring of the main code.

Figure 5 schematizes the data flow, starting from video input to the generation of final reports.

Figure 5 – Data flow and processing pipeline



Source: The author.

Processing occurs in four main stages, described below:

- **Input:** The input layer processes raw video files (`.mp4`), responsible for decoding the video stream and converting frames into processable matrix representations;
- **Detection and Tracking:** In this stage, the YOLOv8n neural network integrated with the BoT-SORT algorithm is applied. Its function is to map the spatial coordinates (x, y) of players and maintain the identity consistency (ID) of each athlete over time, mitigating failures arising from occlusions or camera movement;
- **Team Classification:** Once coordinates are defined, the K-Means algorithm is executed after converting regions of interest into the CIELAB color space. The objective is to group players by the chromatic similarity of uniforms and filter out irrelevant elements (noise and referees);
- **Tactical Inference:** The final module receives the classified positional data and applies sectoral segmentation logic. Based on the statistical distribution of players

on the pitch, the algorithm calculates defense, midfield, and attack lines to infer the predominant tactical formation at the analyzed instant.

As a result, the application generates two artifacts: the rendered video with visual annotations and tactical lines, and a structured dataset containing the frequencies of the respective detected formations.

3.2 DATA ACQUISITION AND MODEL TRAINING

For the object detection component, the YOLOv8 Nano architecture was selected [17]. This choice is justified by two determining factors: first, the need to align the project with the premise of low cost and technological democratization established in this work's objectives; and second, the computational resource constraints available in the development environment (Google Colab, free tier), which would render the training of larger models (such as Large or Extra-Large versions) infeasible in a timely manner.

The dataset used for training was obtained through the computer vision platform Roboflow [29]. The public dataset named "Soccer Player Detection 3" [30], authored by LightWing, composed of 9068 annotated images, was utilized. The database preserved its original split, with 77% of samples allocated to training, 12% to validation, and 11% to testing, ensuring the statistical integrity of the model evaluation.

Training was performed using the Transfer Learning technique, leveraging the network's pre-trained weights to accelerate convergence. It is worth noting that, although the dataset contains annotations for four distinct classes (ball, goalkeeper, player, and referee) and the model was trained to recognize all these objects, the application logic developed in this work implements selective filtering at inference time. The system was programmed to exclusively process detections with the Player label (ID 2), discarding the other classes in post-processing to isolate the object of interest for tactical analysis.

The computational process totaled approximately 2.5 hours on GPU, resulting in a model (file best.pt) optimized for efficient execution on conventional computers.

3.3 TEAM CLASSIFICATION ALGORITHM

Following spatial player detection, the system initiates the semantic association process to determine to which team each individual belongs. This module operates on the Regions of Interest (ROIs) extracted by the YOLO detector.

3.3.1 Color Space and Feature Extraction

In contrast to approaches operating in the RGB (Red-Green-Blue) color space, this work adopted the CIELAB space. The choice is grounded in the search for perceptual uniformity, a central objective of this model's normalization [31].

In this space, the Euclidean distance calculated between two colors was designed to correlate linearly with the visual difference perceived by the human eye. This characteristic gives the system greater robustness against lighting variations (such as stadium shadows), overcoming metric distortions frequent in direct RGB analysis.

For the extraction of each athlete's chromatic feature, a geometric cropping heuristic was implemented. The algorithm analyzes exclusively the central 40% of the detected bounding box height. This technique isolates the torso region, eliminating visual noise coming from the pitch or shorts, and applies the K-Means algorithm ($k = 1$) on the cropped ROI to extract the exact dominant color of that player.

3.3.2 Supervised Initialization and Classification

Unlike fully automatic methods, this system adopts an initial manual configuration to ensure greater precision. Before video processing, the user defines the reference colors of the uniforms (C_{refA} and C_{refB}) using the graphical interface developed with the Streamlit library [32]. This step eliminates the need for the algorithm to "guess" the teams, reducing initialization errors.

With reference colors defined, the system classifies each detected player (P_i) by comparing the color extracted from their torso (C_{P_i}) with the colors of both teams. The decision is based on the minimum Euclidean distance (d) found, according to Equation 1:

$$\text{Team}(P_i) = \text{arg}, \underset{k\in A,B}{\text{min}}, d(C_{P_i}, C_{ref_k}) \tag{1}$$

That is, the algorithm calculates the color difference between the player's uniform and the colors of teams A and B, assigning the player to the team whose color is mathematically closer.

3.3.3 Cardinality-Based Outlier Suppression

An inherent challenge in sports computer vision is the presence of goalkeepers and referees, whose uniforms differ from outfield players. To mitigate erroneous classifications, a filtering algorithm based on maximum team cardinality was developed.

The system monitors the count of players assigned to each team. Given that a team possesses at most 10 outfield players, whenever the system detects $N > 10$ for the same class, a suppression filter is activated. The algorithm identifies, within the excess group, which detection has the greatest Euclidean distance relative to that team's reference color. This element, statistically the least similar to the uniform pattern (likely the goalkeeper), is treated as an outlier and removed from the tactical analysis of that frame.

3.4 TRACKING STRATEGY AND TEMPORAL STABILIZATION

The tracking stage is responsible for maintaining the identity consistency (ID) of each player over time.

3.4.1 Tracking with Motion Compensation

For the execution of this stage, the BoT-SORT algorithm [21] was selected. The adoption of this specific architecture is grounded in preliminary tests performed, where trackers based purely on Kalman Filters (such as DeepSORT [19]) exhibited identity instability due to constant camera movement (pan and zoom) in television broadcasts, frequently resulting in tracking loss due to inconsistent motion predictions.

Thus, the tracking module was configured to utilize BoT-SORT's Camera Motion Compensation (CMC). This functionality aligns adjacent frames prior to motion prediction, allowing the system to isolate the athletes' actual displacement relative to the pitch, ensuring the necessary robustness for ID maintenance even in high-occlusion scenarios.

3.4.2 Stabilization via Voting Window

To mitigate instability in color classification (Team A or Team B), a post-processing step was implemented. The system stores a history of the last 15 team classifications for each active ID. The definition of the player's class at instant t occurs through the calculation of the statistical mode (Mo) over this temporal buffer, according to Equation 2:

$$\text{Final\_Class}(ID_i) = \text{Mo}(\{\text{Class}_t, \text{Class}_{t-1}, \ldots, \text{Class}_{t-14}\}) \tag{2}$$

This heuristic acts as a stability filter, eliminating momentary classification noise (flickering) and ensuring the team's visual cohesion throughout the trajectory.

The choice of the mode is justified by the categorical nature of the data, where arithmetic means are not applicable. The window sizing of 15 frames was defined through experimental tuning. During calibration tests, it was empirically observed that this value provided the best balance for visual stability, being sufficient to suppress classification noise without causing perceptible delays in player identity updates.

3.5 MATHEMATICAL MODELING OF TACTICAL ANALYSIS

The concluding processing phase converts raw spatial coordinates into tactical organization data. Unlike traditional approaches that project fixed zones onto the screen (such as a static grid), the proposed methodology defines game sectors relative to athlete positioning. Thus, the lines delimiting the defense, midfield, and attack are calculated statistically at each instant, compensating for camera framing variations and automatically adapting to team compactness.

3.5.1 Segmentation via Descriptive Statistics

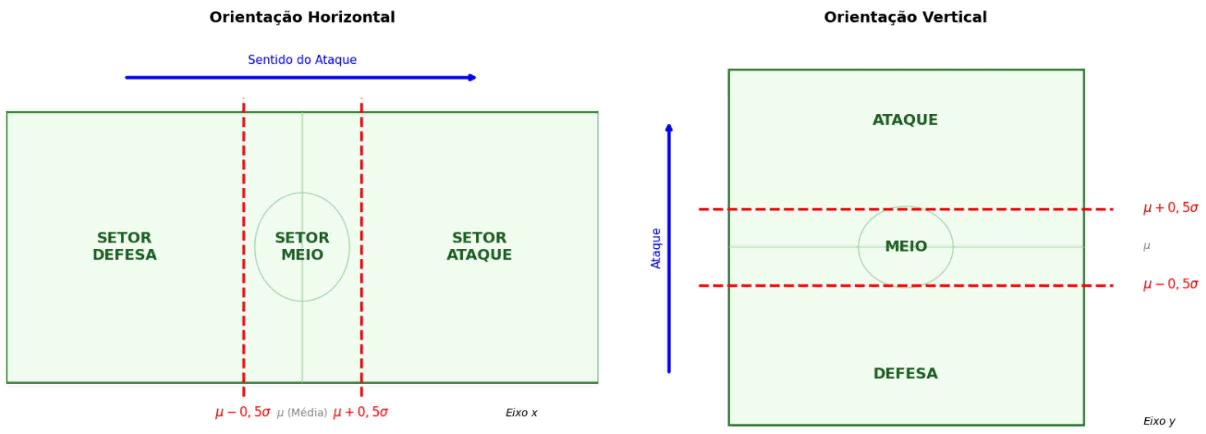
Positional categorization (Defense, Midfield, Attack) is based on the distribution of athletes along the longitudinal axis of the pitch. The algorithm calculates two metrics on the coordinate vector of the outfield players: the arithmetic mean (μ), which establishes the team's geometric center, and the standard deviation (σ), used to measure the degree of dispersion.

The thresholds delimiting the sectors are floating. The classification of a player P_i with coordinate x_i follows the decision rules expressed in Equation 3:

$$\text{Sector}(P_i) = \begin{cases} \text{Defense}, & \text{if } x_i < (\mu - 0,5\sigma) \\ \text{Attack}, & \text{if } x_i > (\mu + 0,5\sigma) \\ \text{Midfield}, & \text{otherwise} \end{cases} \tag{3}$$

It is noteworthy that the implementation inverts the relational operators ($<$ and $>$) according to the configured attack direction (Left $\rightarrow$ Right or vice-versa), maintaining classification consistency regardless of the side of the pitch. The adaptation to different filming orientations (X or Y axes) is conceptually illustrated in Figure 6.

Figure 6 – Adaptation of the statistical model according to camera orientation



Source: The author.

3.5.2 Visualization and Formation Inference

With athletes classified, the system executes grouping by sector. For the visual representation of lines, spatial sorting on the transversal axis is used. The software draws line segments connecting adjacent players in the sorted list, generating the visual effect of tactical lines.

The identification of the tactical formation (e.g., "4-4-2") adopts a deterministic approach based on cardinality. The algorithm counts the number of elements present in each subgroup (defense, midfield, and attack) at the processed instant, inferring the tactical structure without the need for additional predictive models.

3.6 VALIDATION METHODOLOGY

System validation was conducted using a testing protocol aiming to simulate the heterogeneous video conditions frequently found in real sports analyses. This protocol ensures the scientific reproducibility of the presented results.

3.6.1 Environment Configuration and Tools

The prototype was fully implemented in Python, using the Ultralytics library as the basis for the YOLOv8n model [17] and the Streamlit framework [32] to create the user interface. Validation experiments were conducted in a hardware environment emulating the capabilities of conventional analysis machines.

The system input configuration for the experiments included the following manual variables, essential for the supervised classification stage:

- **Reference Colors:** Definition of average uniform colors for Teams A and B in CIELAB space, according to the supervised initialization method detailed in Section 3.4;
- **Pitch Orientation and Attack Direction:** Configuration of the longitudinal direction (Vertical or Horizontal) and attack direction (defining the inversion of relational operators) for the correct application of the statistical model (Section 3.6).

3.6.2 Sample Selection and Characterization

The definition of the test video set was conducted intentionally and strategically. Instead of focusing on a massive quantity of matches, priority was given to situational variety, selecting clips representing the most common and difficult visual challenges in computer vision applied to soccer.

Videos were obtained from TV broadcasts and publicly available tactical recordings. This choice ensures the system is tested under real-world usage conditions, facing the same difficulties an analyst would encounter on a day-to-day basis, rather than solely in controlled environments.

The duration of each sample was standardized to 30 seconds. This delimitation was necessary due to the computational cost of the tracking process. As the algorithm analyzes video frame by frame with motion compensation, processing time for longer files would render tests unfeasible on the available hardware infrastructure. Nevertheless, this interval proved sufficient to capture detection stability and tactical line formation without compromising experimental agility.

Three footage types were defined for validation, as detailed in Table 3:

Table 3 – Video Sampling (Test Set)

Scenario	Description	Main Challenge	Duration
Scenario A	Rear tactical camera (Euro 2020 - Fra vs Sui)	Perspective distortion and players at depth (Feature Extraction Test).	30 s
Scenario B	Standard Television Broadcast (La Liga 2010/11 - Bar vs Real)	Extreme lighting variation and camera movement (Critical CMC Test).	30 s
Scenario C	Fixed Lateral Tactical Camera (La Liga 2025/26 - Real vs Bar)	Lateral occlusions and close-up segmentation precision test.	30 s

Source: The author.

3.6.3 Validation Method

The methodology validation was grounded in a strategy focused on the final product of the tactical analysis. Considering that the work's primary objective is to deliver the reading of the team's organization, tests concentrated on verifying the accuracy of this inference in real-world scenarios.

The evaluation was divided into two levels:

- **Qualitative Perception Assessment:** A structured visual inspection of the detection and tracking algorithms' behavior across the three scenarios. The objective is to identify stability patterns and limitations in stress situations (rapid movement and occlusions), qualifying the system's robustness without the need for granular failure quantification;

- **Tactical Inference Validation (Quantitative):** The objective assessment of the system's capacity to identify the team's predominant formation. The system processed the clips in their entirety, and the detected formation was compared against the official strategy (coach's lineup) to determine the result's convergence.

4

4 RESULTS AND DISCUSSION

This chapter details the experimental validation of the computer vision system. While Chapter 3 established the engineering architecture and algorithmic decisions, this section focuses on demonstrating, quantitatively and qualitatively, the solution's performance.

The evaluation was structured to verify the achievement of the work's objectives under three fundamental pillars:

- **Perception Accuracy:** The system's robustness in correctly detecting, tracking, and classifying outfield players, ensuring precise team segregation;
- **Tactical Inference Fidelity:** The mathematical model's accuracy in identifying instantaneous (momentary) tactical formations, attesting to the validity of the statistical segmentation;
- **Practical Utility:** The system's potential to provide actionable information quickly and intuitively, optimizing the performance analyst's workflow within a low-cost solution context.

In this sense, the discussion of results is organized as follows: Section 4.1 presents the Qualitative Analysis of Visual Perception, discussing the stability of detection and tracking algorithms in test scenarios. Section 4.2 is dedicated to Tactical Inference Validation, comparing the formations identified by the system with the teams' nominal strategies. Finally, Section 4.3 addresses the Study Limitations and observed technical constraints, pointing out directions for future work.

4.1 QUALITATIVE ANALYSIS OF VISUAL PERCEPTION

Prior to assessing the final tactical formation, a qualitative analysis of the performance of the perception subsystems (detection and tracking) was conducted to understand how video and uniform characteristics influence system stability.

In **Scenario A** (Rear Tactical Camera – Euro 2020), a significant performance disparity between the teams was observed. The system demonstrated high visual stability for the French team (blue uniform), maintaining ID consistency and color classification for most of the time, a factor favored by the tactical camera's stability. Conversely, critical failures were observed in detecting the Swiss team (white uniform). As evidenced in Table 4, the quantity of valid detections for this team was insufficient to feed the statistical model, rendering the inference of its tactical formation impossible.

In **Scenario B** (Broadcast with Zoom – La Liga 2010/11), the BoT-SORT algorithm was subjected to the motion compensation test. Visually, the tracker's robustness in maintaining bounding boxes over detected targets was confirmed, even during rapid lateral camera movements (pan). However, selective degradation was noted again: the detection of the Real Madrid team (white uniform) was visibly inferior to that of Barcelona, generating insufficient data for the tactical inference of the Madrid team. For the Barcelona team, formation inference occurred, but oscillations were observed in the assignment of tactical roles (Defense/Midfield/Attack). These variations suggest a sensitivity of the positional categorization algorithm to spatial coordinate instabilities, a phenomenon appearing to be accentuated by the constant movement of the broadcast camera.

Finally, **Scenario C** (Lateral with Shadows – La Liga 2024/25) presented the highest overall perception effectiveness. Unlike the previous cases, the system succeeded in detecting and tracking both teams, including the one with the white uniform, allowing for bilateral tactical inference. However, formation validation in this scenario faced challenges of a different nature: the players' positional dynamicity and the sample's restricted time window hindered convergence toward the static nominal formation, resulting in a reading that reflects the momentary transition more than the team's standard structure.

4.2 TACTICAL INFERENCE VALIDATION

Following the perception analysis, this section is dedicated to the validation of the dynamic sectoral segmentation model (Section 3.6). The objective is to verify whether the application of statistical decision rules ($\mu \pm 0.5\sigma$) on player coordinates results in the correct identification of the team's organizational structure.

For this experiment, the predominant tactical formation (the configuration detected with the highest temporal frequency during the 30-second clip) was compared with the nominal strategy defined by the coaching staff for the respective match. Official lineups were obtained from ESPN and Transfermarkt databases. Table 4 presents the comparative results.

Table 4 – Comparison of formations inferred by the system

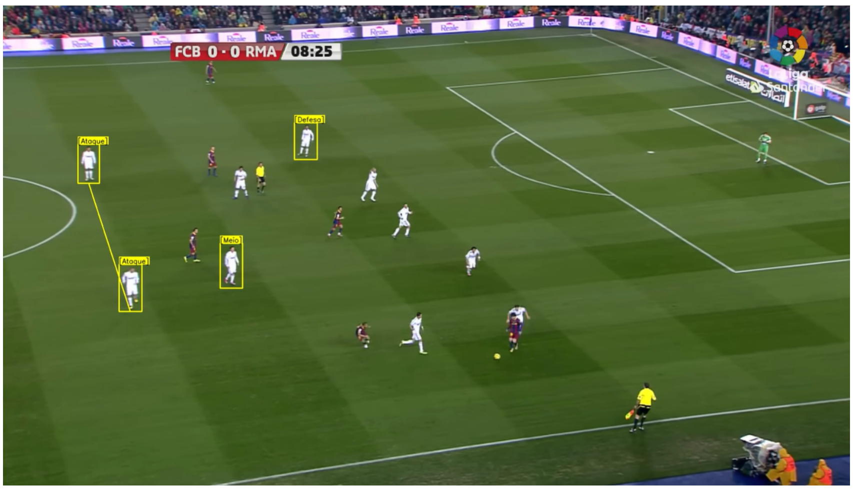
Scenario	Team	Nominal	Inferred (Freq.)	Correspondence
A	France (Euro 2020)	3-4-1-2	3-3-4 (19.4%)	Compatible
A	Switzerland (Euro 2020)	3-4-1-2	Insuff. Detection	Insufficient
B	Barcelona (2011)	4-3-3	4-3-3 (13.1%)	Exact
B	Real Madrid (2011)	4-2-3-1	Insuff. Detection	Insufficient
C	Barcelona (2025)	4-2-3-1	4-2-3 (18.8%)	Incomplete
C	Real Madrid (2025)	4-1-4-1	4-3-3 (28.2%)	Tactical Variation

Source: The author with data from ESPN [33, 34] and Transfermarkt [35].

The analysis of results requires a contextualized interpretation of the nature of soccer and the challenges of computer vision. In cases where there was exact correspondence, such as in Scenario B (Barcelona 2011), the system demonstrated high fidelity in identifying the classic 4-3-3.

However, in the same scenario, the Real Madrid team presented insufficient results. Visual analysis suggests that the low contrast between the white uniform and the bright elements of the environment may have hindered feature extraction by the detector. It was observed that, in moments of higher light incidence, the athletes' edges become less distinguishable, which is consistent with detection failures in low dynamic range scenarios, as illustrated in Figure 7.

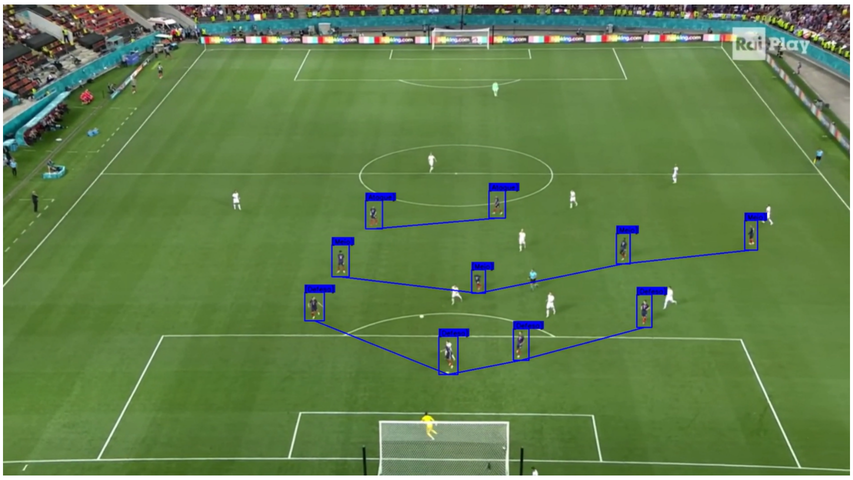
Figure 7 – Visualization of Real Madrid's inconclusive detection (Scenario B)



Source: The author, based on La Liga broadcast.

A specific point of attention refers to the interpretation of formations with four functional lines, such as France's 3-4-1-2 in Scenario A. Figure 8 demonstrates that the algorithm correctly identified the defensive base of three defenders. The variation occurred in the offensive distribution. The predominant inference of a 3-3-4 suggests that the advanced positioning of the wing-backs and the attacking midfielder exceeded the statistical threshold of the midfield sector, being interpreted by the algorithm as presence in the attack sector. This behavior preserves the team's base structure, justifying the classification of the result as compatible.

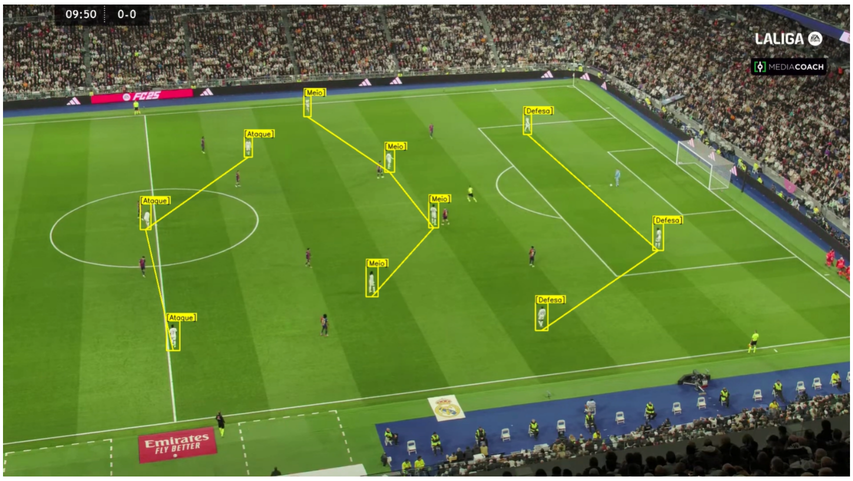
Figure 8 – Detection of tactical formation in defensive moment (Scenario A)



Source: The author, based on RaiPlay broadcast.

Finally, Scenario C (Real Madrid 2025) illustrates the system's capacity to reveal game dynamics. Although the nominal formation was a 4-1-4-1, the system captured the team's offensive reorganization. The inference of a 4-3-3 (28.2% frequency) reflects the moment where wide midfielders push to the attack line and the defensive midfielder aligns with the central midfielders. Figure 9 displays this transition, with the advanced occupation of the lateral corridors.

Figure 9 – Detection of Real Madrid’s offensive dynamics (Scenario C)



Source: The author, based on La Liga broadcast.

This discrepancy reinforces the tool’s practical utility for performance analysis. The system recorded the athletes effective positioning during the offensive phase, highlighting the difference between the dynamic organization on the pitch and the static nominal formation disclosed in the match sheet.

4.3 STUDY LIMITATIONS AND TECHNICAL CONSTRAINTS

The experimental validation confirmed the prototype’s functional viability for automated tactical analysis. However, tests also evidenced technical bottlenecks stemming from design choices aimed at low cost and hardware constraints. This section discusses the main challenges faced in the current implementation.

4.3.1 Sensitivity of the YOLOv8 Nano Architecture

The main qualitative limitation observed in the results stems from the choice of the YOLOv8 Nano model [17]. Although this version is optimized for execution in resource-constrained environments (such as cloud service free tiers), its reduced parameter density imposes a trade-off between efficiency and sensitivity (Recall).

This sensitivity restriction manifested critically regarding contrast and luminance. In the conducted experiments, specifically in Scenarios A and B, a degradation was noted in the detection rate of players with white uniforms when exposed to conditions of high solar incidence. This phenomenon is consistent with the loss of texture and edge information caused by pixel saturation in broadcast images. The implemented pipeline, by processing frames directly without histogram equalization or local contrast enhancement steps, became susceptible to this visual camouflage, hindering feature extraction by the model in high-luminosity zones.

It was also observed that, in wide shot scenarios (Scenario A), the instability of the detected tactical formations was not caused by a failure in the statistical logic, but rather by the primary non-detection of distant players. The Nano model tends to ignore small objects (few pixels in height), generating momentary gaps in data collection that affect the calculation of the team's positional mean. The use of more robust architectures, such as YOLOv8 Large, would mitigate this problem, although it would require dedicated hardware incompatible with this work's accessibility premise.

4.3.2 Processing Latency and Temporal Viability

Although the system operates with a lightweight detection model, the computational cost of tracking proved to be a limiting factor. The combination of the BoT-SORT algorithm (which performs frame-by-frame image alignment for motion compensation) with neural network inference generated significant accumulated latency.

In the tests performed, the processing time for short 30-second videos exceeded real-time playback time, highlighting that the solution, on the current infrastructure, is strictly oriented towards post-match analysis (offline). To enable the tool's use during the match (live tagging), migration to an environment with high-performance local GPU acceleration would be mandatory, eliminating the processing bottleneck that currently limits the experience on conventional personal computers.

5

5 CONCLUSION

This work addressed the challenge of technological accessibility in sports performance analysis, proposing an engineering solution capable of automating the extraction of tactical intelligence from television broadcast videos. The research's central motivation was grounded in the premise that the application of modern Computer Vision techniques, when optimized for resource-constrained scenarios, can act as a vector for democratization in soccer, reducing the technical gap between elite clubs and teams with lower investment capacity.

Revisiting the general objective established at the beginning of this study, it is concluded that the goal of developing and validating a methodology for mapping tactical formations was achieved. The constructed computational artifact demonstrated technical viability by processing unstructured video streams and converting them, autonomously, into structured representations of the teams' spatial organization. The successful integration of detection, tracking, and statistical modeling algorithms proved that it is possible to extract high-level metrics without mandatory reliance on proprietary hardware or multi-million dollar optical tracking systems.

However, experimental validation also revealed that the pursuit of low cost imposes technical trade-offs. The choice of lightweight models and the reliance on color heuristics restrict the current solution's applicability to controlled lighting conditions and require human supervision during initial configuration, characterizing the system as a semi-automatic support tool rather than a fully autonomous one.

5.1 SYNTHESIS OF CONTRIBUTIONS

The primary academic and technical contribution of this work lies in the architecture of the proposed processing pipeline. The combination, in this specific context, of the BoT-SORT tracker with the voting window proved to be a robust response to the challenges imposed by camera movement in TV broadcasts. Where conventional trackers failed due to the loss of linearity in motion, the developed solution maintained the cohesion of visual identities, proving that camera motion compensation is an indispensable technical condition for broadcast-based tactical analysis software.

Another relevant contribution was the implementation of domain heuristics for visual data refinement. The application of cardinality filters (limiting detection to 10 outfield players) and outlier suppression based on chromatic distance proved essential to mitigate failures inherent to detection models, eliminating noise such as goalkeepers and referees. This hybrid approach, combining AI probability with the sport's deterministic rules, was decisive in ensuring tactical data integrity.

Finally, the mathematical condition of dynamic sectoral segmentation presented itself as an effective alternative to static grid methods. By utilizing the statistical dispersion (standard deviation) of the athletes themselves to define defense and attack lines, the algorithm endowed the system with an adaptive capacity, allowing for correct tactical interpretation regardless of the team's compactness level or the zoom applied by the broadcast.

5.2 FUTURE WORK

The development of this study has opened new avenues of investigation for the enhancement of automated tactical analysis. As an immediate evolution, the replacement of the YOLOv8 Nano detection model with larger architectures, such as the Large versions or Transformer-based models, is suggested. This aims to mitigate detection failures in wide shots, even if it demands migration to dedicated cloud processing infrastructures.

In the future, the application of perspective transformation algorithms to generate a top-down view of the pitch presents itself as the logical next step. The ability to project player coordinates onto a two-dimensional map would allow not only for more precise positional analysis but also enable the calculation of advanced metrics such as Pitch Control, heatmaps, and passing networks. Finally, the integration of Re-Identification (ReID) models based on deep features, replacing color clustering, would drastically reduce the dependence on manual inputs, paving the way for a fully autonomous tactical analysis system resilient to lighting variations.

REFERENCES

REFERENCES

[1] C. Carling, A. M. Williams, and T. Reilly, *Handbook of soccer match analysis: a systematic approach to improving performance*. London: Routledge, 2005.

[2] M. Lewis, *Moneyball: the art of winning an unfair game*. New York: W. W. Norton & Company, 2004.

[3] M. D. Hughes *et al.*, “Moneyball and soccer: an analysis of the key performance indicators of elite male soccer players by position,” *Journal of Human Sport and Exercise*, vol. 7, no. 2, pp. 402–412, 2012.

[4] X. Xiao, A.-K. Kupfer, I. Benbasat, and R. T. Cenfetelli, “Sports digitalization: An overview and a research agenda,” in *Proceedings of the 38th International Conference on Information Systems (ICIS)*. Seoul: Association for Information Systems, 2017, pp. 1–19.

[5] Nacsport, “Nacsport pricing plans,” https://www.nacsport.com/pricing, 2025, orem, UT. Accessed: 9 June, 2025.

[6] Hudl, “Sportscode tiers,” https://www.hudl.com/products/sportscode/tiers, 2025, lincoln, NE. Accessed: 9 June, 2025.

[7] ——, “Wyscout pricing,” https://www.hudl.com/products/wyscout/pricing, 2025, lincoln, NE. Accessed: 9 June, 2025.

[8] Catapult, “Pricing,” https://www.catapult.com/pricing#buynow, 2025, melbourne, AU. Accessed: 9 June, 2025.

[9] StatsBomb, “The next generation of football data,” https://statsbomb.com/, 2025, bath, UK. Accessed: 9 June, 2025.

[10] Metrica Sports, “Plans,” https://www.metrica-sports.com/pt/plans, 2025, netherlands. Accessed: 9 June, 2025.

[11] GloboEsporte, “Veja cotas financeiras da série d de 2025: saiba os valores e o calendário de pagamento aos clubes,” 2025, accessed: 11 Dec. 2025. [Online]. Available: https://ge.globo.com/pi/futebol/brasileirao-serie-d/noticia/2025/04/11/veja-cotas-financeiras-da-serie-d-de-2025-saiba-os-valores-e-o-calendario-de-pagamento-ghtml

[12] M. E. d. P. Shamah, “Análise de desempenho no futebol: a prática do analista de desempenho nas categorias de base dos clubes brasileiros da série a,” Master’s Thesis, Universidade Federal do Rio Grande do Sul, Escola de Educação Física, Fisioterapia e Dança, Porto Alegre, 2021.

[13] R. Groom, C. Cushion, and L. Nelson, "The delivery of video-based performance analysis by england youth soccer coaches: Towards a grounded theory," *Journal of Applied Sport Psychology*, vol. 23, no. 1, pp. 16–32, 2011.

[14] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews, "Discovering team structures in soccer from spatiotemporal data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2596–2610, out 2016.

[15] F. R. Goes *et al.*, "Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review," *European Journal of Sport Science*, vol. 21, no. 4, pp. 481–496, 2021.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016, pp. 779–788.

[17] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," 2023, accessed: 22 Nov. 2025. [Online]. Available: https://github.com/ultralytics/ultralytics

[18] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[19] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Beijing: IEEE, 2017, pp. 3645–3649.

[20] G. Welch and G. Bishop, "An introduction to the kalman filter," Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, Technical Report TR 95-041, 1995.

[21] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," *arXiv preprint arXiv:2206.14651*, 2022. [Online]. Available: https://arxiv.org/abs/2206.14651

[22] M. Rico-González *et al.*, "Machine learning application in soccer: a systematic review," *Biology of Sport*, vol. 40, no. 1, pp. 249–263, 2023.

[23] F. Zheng, D. Z. Al-Hamid, P. H. J. Chong, C. Yang, and X. J. Li, "A review of computer vision technology for football videos," *Information*, vol. 16, no. 5, p. 355, 2025.

[24] K. Kausalya, K. S. S. Raja, and S. Sudha, "An optimized deep learning approach for detection and classification of player actions in football game," *Entertainment Computing*, vol. 55, p. 101003, 2025.

[25] A. Cioppa, S. Giancola, A. Deliège, L. Kang, X. Zhou, Z. Cheng, B. Ghanem, and M. Van Droogenbroeck, "SoccerNet-Tracking: A large-scale dataset and benchmark for multiple object tracking in soccer videos," in *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports (MMSports'22)*. ACM, 2022, pp. 81–91.

[26] R. Theagarajan and B. Bhanu, "An automated system for generating tactical performance statistics for individual soccer players from videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 632–644, 2021.

[27] W. Huang, S. Wang, and P. Li, "The application of deep learning in tactical analysis of football matches," *Research Square*, 2025, preprint.

[28] R. S. Wazlawick, *Metodologia de Pesquisa para Ciência da Computação*, 2nd ed. Rio de Janeiro: Elsevier, 2014.

[29] Roboflow Inc., "Roboflow: Give your software the sense of sight," https://roboflow.com, 2025, software. Accessed: 22 Nov. 2025.

[30] LightWing, "Soccer player detection 3 dataset," https://universe.roboflow.com/lightwing/soccer-player-detection-3, Jan. 2023, accessed: 22 Nov. 2025.

[31] Commission Internationale de l'Eclairage, "Colorimetry," CIE Central Bureau, Vienna, Publication 15:2004, 2004, define o espaço L*a*b* e suas propriedades de uniformidade perceptual.

[32] Streamlit Inc., "Streamlit: A faster way to build and share data apps," https://streamlit.io, 2025, software. Accessed: 22 Nov. 2025.

[33] ESPN, "França 3-3 suíça (28 de jun, 2021) resumo do jogo," https://www.espn.com.br/futebol/partida/_/jogoId/560303/suica-franca, 2021, accessed: 22 Nov. 2025.

[34] ——, "Real madrid vs barcelona: Resumo da partida," https://www.espn.com.br/futebol/partida/_/jogoId/748236/barcelona-real-madrid, 2025, accessed: 22 Nov. 2025.

[35] Transfermarkt, "Fc barcelona - real madrid cf: Ficha de jogo 29/11/2010," http://transfermarkt.com.br/fc-barcelona_real-madrid-cf/index/spielbericht/1036195, 2010, accessed: 22 Nov. 2025.

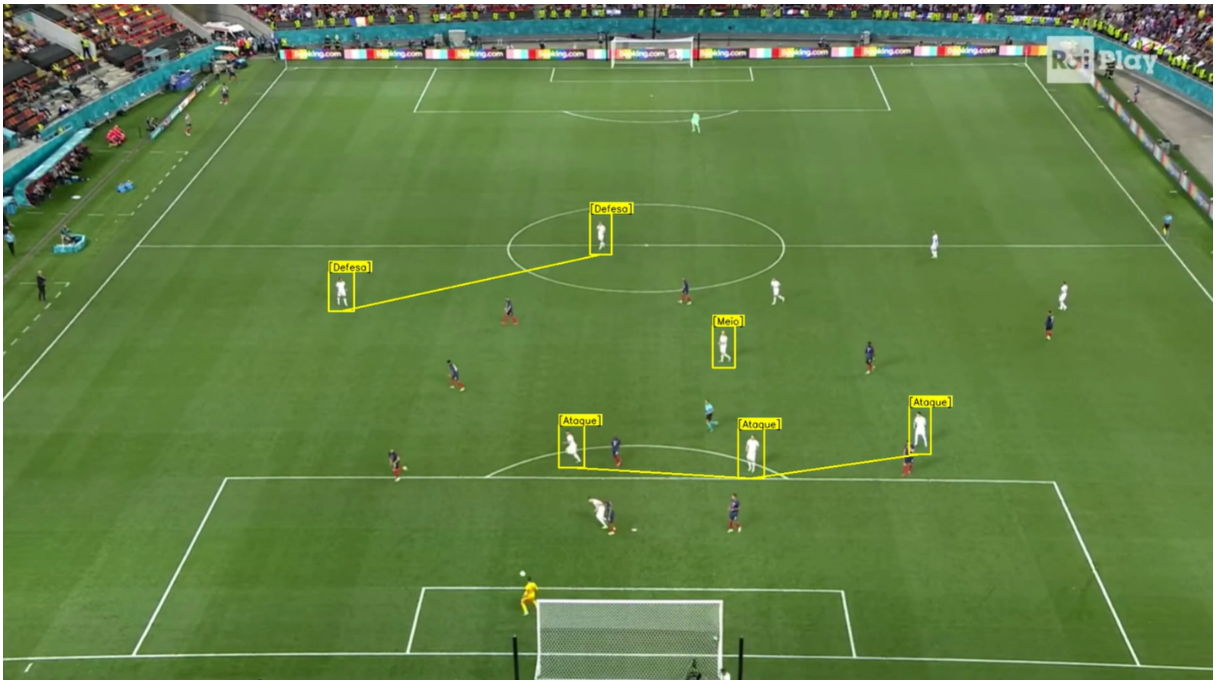
APPENDICES

Appendix A - Interface and Tests

Figure 10 – Interface

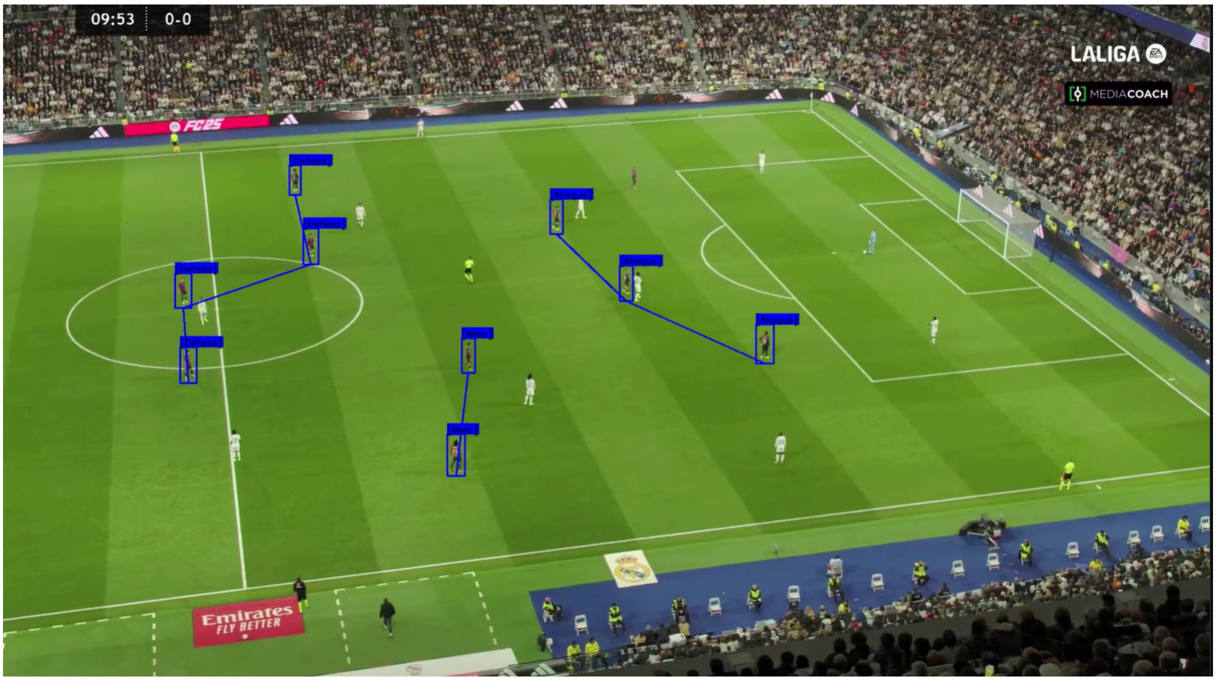
Source: The author.

Figure 11 – Switzerland detection (Scenario A)



Source: The author.

Figure 12 – Barcelona detection (Scenario C)



Source: The author.

Appendix B - Interview

Participant: Performance Analyst at Capital FC.
Methodology: Semi-structured interview via audio recordings.
Collection date: May 29, 2025

1. What are the department's main needs and deadlines?

The analyst reported a scarcity of information available regarding the Série D and the local championship (Candangão). The restriction of human resources was highlighted, with only one professional in the department, creating difficulties for the manual search of information. The immediate need for technologies that facilitate player scouting was expressed to optimize work time.

2. How do current data collection processes work?

Currently, there is no data collection process focused on tactical performance. The only use of structured data is restricted to the physiological aspect (load and injury prevention) via GPS. It was reported that previous attempts at technological implementation to aid recruitment failed due to low reliability in the available data.

3. What are the tactical analysis method and tools used?

The *Nacsport* software is used, although the analyst is knowledgeable about solutions such as *Hudl*. The routine is based on watching three opponent matches to identify behavioral patterns (home, away, and after substitutions). The process relies heavily on the manual tagging function to detect offensive and defensive patterns, and set-pieces.

4. Which metrics (KPIs) are used for evaluation?

In the individual scope, analysis is based on the player's position to identify improvements in positioning and reaction. In the collective scope, specific statistics demanded by the coach are used, highlighting the frequency of team entries into the final third of the pitch and the identification of the corridors where plays occur.


SEJAM BEM-VINDOS
idp
Ensino que
te conecta