

UAB

Universitat Autònoma de Barcelona

Breaking the Data Deadlock

An End-to-End AI Pipeline to Introduce Data Culture in Futsal

Bachelor's Final Thesis (Treball de Fi de Grau)

Degree: Business and Technology (Empresa i Tecnologia)

Author: Isaac Boqué Carbonell

Supervisor: Xavi Verge Mestre

Academic Year 2025–2026

Submission Date: 26th May 2026

Acknowledgements

To Xavi Verge Mestre, my supervisor, whose flexibility and guidance were essential throughout this project. He consistently provided direction while respecting my decisions and the path I chose to take in the research - a balance rarer, and more valuable, than it may first appear.

To my family, who have supported me unreservedly at every step of my life. Their unwavering confidence in me has been a constant source of motivation and strength.

To Judit, who generously helped me gather information and introduced me to tools and resources I did not even know existed. She was always willing to share her knowledge, while remaining equally open to learning from others - even from “just an intern”.

And finally, to all the years I have spent in the world of futsal. Those years gave me something no library could provide: the network, the credibility, and the access to professionals who shared privileged data and genuine insights that became the empirical backbone of this work.

Abstract

ENGLISH ABSTRACT

Amateur and semi-professional futsal clubs across Spain operate in an analytics vacuum. Despite clear demand from coaches and analysts for performance data, the vast majority of clubs at this level lack the tools, budget, and technical staff to collect, process, and act on structured match data. This thesis diagnoses the root cause of this gap - not cultural resistance, but an absence of accessible and automated data-collection infrastructure - and presents a working solution designed to break the cycle.

The research begins with six in-depth interviews with futsal professionals spanning the full competitive spectrum in Spain, from the national Catalan team to semi-professional clubs in the first regional division. The interviews surface a consistent pattern: every professional knows what metrics they would like to measure, none has a reliable and scalable system to measure them, and those who have built any data practice have done so at extraordinary human cost. The conclusion is unambiguous - the barrier is structural, not attitudinal.

On this empirical foundation, the thesis proposes and builds a Python-based AI pipeline that transforms standard fixed-camera match footage into structured, actionable data without any specialist hardware or dedicated human operators. The pipeline integrates five iteratively developed prototypes: automated YouTube video access via yt-dlp and OpenCV, player and ball detection using YOLOv11n, persistent multi-object tracking with ByteTrack, team classification using K-Means clustering in HSV colour space, and a dual-homography perspective transformation (PiecewiseTransformer) that maps camera-space pixel coordinates to real-world 2D pitch coordinates. A field-boundary filtering layer ensures only in-play positions contribute to KPI accumulation.

The system generates per-player KPIs validated by the interviewed professionals: distance covered, sprint count and duration, duel count, and possession time. These metrics are rendered in real time on a tactical overhead board and exported as structured CSV and JSON datasets for downstream dashboard visualisation.

The thesis is framed and executed using an adapted Scrum methodology, treating each prototype as a sprint with explicit retrospectives and documented design decisions. The result is not a proof-of-concept notebook but a reproducible, documented system that any club with a fixed camera and a laptop can deploy. The broader argument this work makes is both technical and cultural: data culture in futsal is not a luxury or a distant aspiration. It is achievable now, with existing video infrastructure, at a cost any semi-professional club can afford.

RESUM (CATALAN)

Els clubs de futbol sala amateur i semi-professional operen en un buit analític. Tot i la demanda clara de dades de rendiment per part d'entrenadors i analistes, la gran majoria dels clubs d'aquest nivell no disposen de les eines, el pressupost ni el personal tècnic per recollir, processar i actuar sobre dades estructurades de partit. Aquesta tesi diagnostica la causa arrel d'aquest buit - no la resistència cultural, sinó l'absència d'infraestructura de recollida de dades accessible i automatitzada - i presenta una solució funcional dissenyada per trencar el cicle.

KEYWORDS / PARAULES CLAU

Futsal analytics · Computer vision · Player tracking · YOLO · ByteTrack · Homography · KPI · Data culture · Sports technology · Agile development

List of Abbreviations and Symbols

Abbreviation	Full Form
AI	Artificial Intelligence
CV	Computer Vision
DL	Deep Learning
FCF	Federació Catalana de Futbol
fps	Frames per second
GPU	Graphics Processing Unit
HSV	Hue, Saturation, Value (colour space)
IoU	Intersection over Union
KPI	Key Performance Indicator
mAP	Mean Average Precision
YOLO	You Only Look Once (object detection model family)
2D	Two-Dimensional
xG	Expected Goals
GDPR	General Data Protection Regulation

Table of Contents

Abstract	3
List of Abbreviations and Symbols	5
1. Introduction	9
1.1. Motivation.....	9
1.2. Research Perspective	9
1.3. Background and Context	10
1.4. Research Hypothesis and Objectives.....	12
1.5. Structure of the Thesis	12
2. State of the Art.....	13
2.1. The Rise of Sports Analytics	13
2.2. Computer Vision in Sports.....	14
2.3. Futsal-Specific Research.....	15
2.5. The Specific Gap	17
3. Methodology.....	18
3.1. Research Design.....	18
3.2. Qualitative Research: Interview Methodology	18
3.3. Technical Development: Adapted Scrum	19
3.4. Data Collection and Ethical Considerations.....	20
3.5. Geographical and Contextual Scope	20
4. Project Definition and Scope.....	22
4.1. Central Objective	22
4.2. Operational Objectives	22
4.3. Project Scope.....	22
4.4. Value Criteria for Each Prototype.....	23
5. KPI Framework for Futsal.....	24
5.1. Why KPIs Matter - and Why They Are Poorly Defined in Futsal	24
5.2. Derivation of the KPI Framework	24
5.3. The Proposed KPI Framework.....	25
6. Problem Diagnosis: The Absence of Data Culture in Futsal	27
6.1. Research Methodology	27
6.2. Interview Findings by Professional	27
Jordi Barrero · National Team Selector, FCF	27
Martí Puig · Analyst, Catalan National Team FCF	27
Roger Bonet · Former Assistant Coach, FC Barcelona Futsal	28
Guillem Katz · Assistant Coach, Hospitalet Bellsport	28
Pol Novo · Head of Coaches, Futsal Mataró	29
Sergi Gil · Assistant Coach, Noia Futsal (First Division).....	29
6.3. The Transversal Pattern.....	29

6.4.	A Self-Reinforcing Deadlock: No Data, No Framework; No Framework, No Data	30
6.5.	Why Existing Solutions Do Not Resolve the Gap	31
7.	Technical Development: The Five Prototype Sprints.....	32
7.1.	System Architecture.....	32
7.2.	Sprint 1 - Automated Video Ingestion	33
7.2.1.	Digital Video as a Spatiotemporal Signal	33
7.2.2.	Streaming Architecture.....	33
7.3.	Sprint 2 - Classical Tracking Approaches: Mathematical Foundations and Failure Analysis.....	34
7.3.1.	Colour-Space Representation: Why HSV	34
7.3.2.	Approach A: Colour-Based Tracking - Formal Model.....	34
7.3.3.	Approach B: Lucas-Kanade Optical Flow - Derivation.....	35
7.3.4.	Outcome.....	36
7.4.	Sprint 3 - YOLO Detection and Dual-Tracker Architecture	36
7.4.1.	YOLO Architecture and mAP.....	36
7.4.2.	PlayerTracker - Greedy Hungarian Algorithm	37
7.4.3.	BallTracker - Kalman Filter	38
7.4.4.	Team Classification - K-Means on HSV.....	39
7.5.	Sprint 4 - Perspective Geometry: 6-Point Homography	40
7.5.1.	The Projective Camera Model.....	40
7.5.2.	Why 6 Points Instead of 4.....	41
7.5.3.	Coordinate Systems	41
7.5.4.	Reprojection Error Analysis	42
7.6.	Sprint 5 - KPI Derivation and Spatial Filtering	42
7.6.1.	KPI Derivation from Positional Data.....	42
7.6.2.	Spatial Filtering - Three-Layer Pipeline	43
7.7.	Data Outputs and Formats	44
7.7.1.	Output Pipeline.....	44
7.7.2.	JSONL Position Schema	45
7.7.3.	KPI CSV Schema	45
7.8.	Complete Pipeline Architecture and Computational Complexity	45
8.	Training YOLO11 on YouTube Futsal Data	47
8.1.	Prerequisites	47
8.2.	Step 1: Extract Frames from YouTube.....	47
8.3.	Step 2: Annotate Frames	48
8.4.	Step 3: Split into Train and Validation Sets	49
8.5.	Step 4: Train YOLO11.....	49
8.6.	Step 5: Deploy the Fine-Tuned Model.....	50
8.7.	Minimum Dataset Size and Training Time Estimates.....	50

8.8.	Tips for Better Results	50
8.9.	Common Issues.....	51
9.	Analysis of Results.....	52
9.1.	Validation of the Qualitative Research	52
9.2.	Technical Pipeline Performance	52
9.3.	Alignment Between KPI Framework and Pipeline Output	53
9.4.	Limitations.....	53
10.	Conclusions.....	55
10.1.	Verification of the Central Hypothesis	55
10.2.	Key Contributions	55
10.3.	Future Work.....	56
10.4.	Unexpected Findings and Personal Reflection.....	57
	References.....	61
	Annex A: Interview Instrument.....	63
	Annex B: Sprint Retrospective Template	63
	Annex C: Full Technology Stack	63
	Annex D: Configuration Parameter Reference	64

1. Introduction

1.1. Motivation

The seed of this thesis was planted on a futsal court. As a semi-professional futsal player with over a decade of competitive experience in Catalonia, I have spent thousands of hours in training sessions and matches where decisions - about tactics, player selection, and physical management - were made almost entirely on intuition and memory. The conversations after matches, the debates in dressing rooms, the selection calls that determined who played and who sat on the bench: all of it was driven by subjective observation.

Then I encountered a sentence that would not leave me alone: "Without data, you are just another person with an opinion." The attribution varies - it is often associated with W. Edwards Deming, the American statistician and management theorist - but the force of the idea is independent of its origin. It named something I had been experiencing for years without having a framework for it. Futsal, at the semi-professional level, runs almost entirely on opinions.

This thesis is, at its core, the attempt to do something about that. Not through abstraction or theoretical modelling, but through a working system that any club with a camera and a laptop could deploy. The motivation is personal, but the problem is structural - and the solution, as this work demonstrates, is within reach.

1.2. Research Perspective

This research is approached from the intersection of three perspectives: that of a practitioner with deep domain knowledge of the sport, that of a business and technology student trained to analyse systems and identify value propositions, and that of a builder

who believes that the most honest way to validate a technical hypothesis is to implement it and test it against real-world data.

The project does not claim to produce the definitive analytics platform for futsal. It claims to demonstrate that the core technical problem - extracting structured positional data from standard fixed-camera video - is solvable with existing open-source tools, at a cost accessible to semi-professional clubs, and to a level of precision sufficient to generate actionable KPIs. That is the claim, and the system built over five iterative prototypes is the evidence.

1.3. Background and Context

The transformation of professional sport by data analytics has been well documented since Michael Lewis's *Moneyball* (2003) brought sabermetrics into mainstream conversation. What Lewis chronicled in baseball - the systematic replacement of gut feel with empirical measurement - has since spread across virtually every major sport. Football clubs employ teams of data scientists. Basketball franchises make roster decisions based on advanced shot-quality metrics. Tennis players review detailed movement heat maps between sets.

Futsal, however, occupies a peculiar position in this landscape. Despite being one of the most technically demanding small-sided sports - with its intense intermittent efforts, rapid transitions, and premium on spatial intelligence - it has received comparatively little attention from the sports analytics research community. The tools and systems that have democratised data in football, basketball, and tennis have not made a corresponding leap into the futsal ecosystem at the semi-professional level.

The reasons are partly economic (smaller budgets, smaller markets) and partly structural (no centralised data infrastructure at lower competitive levels). The result is a data

deadlock: clubs lack the tools to collect data, and because they lack data, they cannot validate which metrics matter, which in turn removes the incentive to invest in collection infrastructure. This thesis identifies breaking that deadlock as its central challenge.

1.4. Research Hypothesis and Objectives

The central hypothesis of this thesis is: an AI pipeline built on standard open-source computer vision tools can transform fixed-camera futsal match footage into structured, validated, actionable performance data at a cost accessible to semi-professional clubs, without specialist hardware or dedicated human operators.

The research pursues four operational objectives:

- (1) To empirically validate the existence and nature of the data gap in semi-professional futsal through primary qualitative research with domain professionals.
- (2) To identify a set of KPIs relevant to futsal performance, validated by those professionals, that can serve as the target output of an automated system.
- (3) To design and build a working AI pipeline that generates those KPIs from standard fixed-camera video, documented in sufficient detail to be reproducible.
- (4) To reflect critically on the limitations of the system, the methodological choices made, and the conditions under which the hypothesis holds.

1.5. Structure of the Thesis

The thesis follows the structure of the research process itself. Chapter 2 presents the state of the art in sports analytics and computer vision. Chapter 3 details the agile methodology adopted. Chapter 4 defines the project scope and success criteria. Chapter 5 presents the KPI framework validated with professionals. Chapter 6 documents the five iterative prototype sprints that form the technical core of the work. The thesis concludes with the analysis of results, conclusions, and a list of references followed by technical annexes.

2. State of the Art

2.1. The Rise of Sports Analytics

The history of data-driven decision-making in sport is, in many ways, the history of overcoming institutional resistance to empirical evidence. Bill James, writing his self-published *Baseball Abstracts* in the early 1980s, was considered an eccentric outsider by the baseball establishment. Two decades later, the methods he pioneered - subsequently labelled *sabermetrics* - were standard practice in Major League Baseball. The pattern has repeated across sports: initial scepticism, a period of competitive advantage for early adopters, and then rapid diffusion as the evidence became undeniable.

The defining popular articulation of this shift, Michael Lewis's *Moneyball* (2003), documented how the Oakland Athletics used statistical analysis to compete against franchises with vastly superior financial resources. The book's central insight - that certain performance metrics were systematically undervalued by the market - has since been generalised into a broader principle of competitive analytics that Thomas Davenport and Jeanne Harris formalised in their influential *Competing on Analytics* (2007): that organisations which build superior analytical capabilities can sustain competitive advantages that are difficult for rivals to replicate.

In football, the transformation has accelerated sharply since the early 2010s. The widespread adoption of expected goals (xG) as a measure of shot quality, pioneered by analysts including Sam Green and later popularised by Opta, fundamentally changed how clubs evaluated attacking and defensive performance. Clubs such as Liverpool FC and Bayer Leverkusen have publicly attributed competitive advantages to their analytics departments. The emergence of optical tracking providers - SkillCorner, Second Spectrum, StatsBomb - has made sub-second positional data for all players available to top-flight clubs.

2.2. Computer Vision in Sports

The application of computer vision to sports analysis has developed in parallel with advances in deep learning, particularly convolutional neural networks. Early systems relied on colour-based segmentation and optical flow to track players and balls in video footage. The limitations of these approaches - sensitivity to lighting variation, inability to distinguish players in close proximity, difficulty maintaining identity through occlusions - constrained their applicability to controlled broadcast environments.

The introduction of deep learning-based object detection, culminating in the YOLO (You Only Look Once) family of models, dramatically improved the reliability and speed of player and ball detection. YOLO's single-pass architecture, which processes an entire frame in one forward pass of the network, made real-time detection feasible on standard hardware. The subsequent development of robust multi-object tracking algorithms - DeepSORT, FairMOT, and ByteTrack - addressed the identity-maintenance problem by combining appearance features with motion prediction, enabling persistent player IDs through occlusions and crossings.

Homography-based perspective transformation - the mathematical procedure that maps camera-space pixel coordinates to real-world pitch coordinates - has been applied in sports analytics since at least the early 2000s. The key challenge at the semi-professional level is that robust homography computation typically assumes a controlled camera setup and consistent calibration. The dual homography approach implemented in this thesis (PiecewiseTransformer) addresses the non-uniform perspective distortion inherent in wide-angle fixed-camera setups, reducing positional error at pitch margins by over 40% compared to a single homography.

2.3. Futsal-Specific Research

The academic literature on futsal performance analytics is sparse relative to football. The existing research has primarily addressed physical and physiological demands - sprint profiles, heart rate responses, and high-intensity effort distributions - rather than tactical or technical metrics. Studies by Caetano et al. (2015) on sprint sequences in professional futsal, and by García-Unanue et al. (2017) on physical demands across playing positions, have established that futsal involves significantly higher ball-contact frequency and shorter recovery intervals than outdoor football, placing unique demands on both physical conditioning and tactical decision-making.

The specific KPIs relevant to futsal coaching - duel success rate, pressing effectiveness, possession chains, and set-piece outcomes - have received very limited empirical treatment. The KPI framework proposed in Chapter 5 of this thesis draws on both the existing literature and the primary research conducted with Spanish futsal professionals, filling a gap that the academic literature has not yet addressed systematically.

2.4. The Technology Frontier: ISE 2026, SAM 3, and the Future of Sports Tracking

In February 2026, as part of the research for this thesis, the author attended the Integrated Systems Europe 2026 (ISE) trade fair in Barcelona, where Sony demonstrated a sports analytics system based on a motorised high-resolution camera capable of autonomously tracking a single subject - either the ball or a designated player - across a basketball court in real time. The demonstration represented the cutting edge of hardware-based tracking solutions for elite sports and broadcast production environments.

However, the Sony system also highlighted why the gap identified in Chapter 1 remains unresolved. The solution depends on dedicated motorised camera hardware rather than

the fixed tripod-mounted camera typically available to semi-professional futsal clubs. Furthermore, it focuses on tracking a single subject rather than capturing the complete tactical context involving all players on the pitch. Its positioning is therefore fundamentally different from the objective of this thesis: Sony is addressing the needs of elite broadcast production, whereas this work focuses on affordable, multi-player, full-pitch tracking using standard video infrastructure accessible to grassroots and semi-professional clubs.

At the same time, the computer vision field is undergoing a broader paradigm shift driven by segmentation-based models. One of the most significant recent developments is SAM 3 by Meta AI (Segment Anything Model 3), released in May 2026. Unlike traditional tracking systems based on bounding boxes - including the pipeline developed in this thesis - SAM 3 produces pixel-accurate segmentation masks for detected entities, enabling object-level understanding of a scene rather than simple region-level localisation.

The implications of this approach for sports analytics are substantial. SAM 3 introduces strong zero-shot generalisation capabilities, allowing accurate segmentation of objects and classes not explicitly seen during training. In the context of this thesis, such capabilities could potentially replace both the YOLO detection stage and the team-colour classification stage. By providing full player silhouettes rather than approximate bounding boxes, segmentation masks would enable direct extraction of jersey colours at pixel level, eliminating the cropping approximations currently required by the K-Means colour-classification pipeline.

Despite its potential, the main limitation preventing immediate integration of SAM 3 into the proposed system is computational cost. Real-time segmentation of a 1080p video stream at 25 frames per second exceeds the capabilities of the standard laptop CPU targeted by this thesis. As a result, although segmentation-based architectures represent

the likely future of sports tracking systems, they remain impractical for the low-cost deployment constraints that define this project.

For this reason, the system developed in this thesis should be understood as a deployable bridge between current practical constraints and a rapidly evolving technological frontier. As GPU hardware becomes more affordable and efficient, integrating segmentation-based approaches such as SAM 3 represents the most important direction for future development of the system.

2.5. The Specific Gap

The review of the state of the art reveals a clear structural gap. High-quality automated sports analytics tools exist, but they are designed and priced for elite institutions. The open-source tooling for computer vision and multi-object tracking has reached a maturity level that makes semi-professional-grade tracking technically feasible. But no integrated, documented, and validated pipeline exists that is specifically designed for the futsal context, deployable without specialist hardware or technical staff, and priced at a point accessible to clubs with limited budgets. That is the gap this thesis addresses.

3. Methodology

3.1. Research Design

The research design integrates two complementary methodological approaches. The first is qualitative primary research: six in-depth semi-structured interviews with futsal professionals, conducted to establish the empirical basis for the problem diagnosis and to validate the KPI framework. The second is iterative technical development: five prototype sprints, each addressing a specific technical challenge in the computer vision pipeline, following an adapted Scrum framework.

This combination reflects the nature of the problem. A purely academic treatment of sports analytics risks producing frameworks that are theoretically coherent but practically unusable. A purely technical project risks building a solution to a problem that has not been properly understood. The integration of qualitative domain research with iterative technical development ensures that the system built is grounded in real professional needs, and that the technical decisions are driven by empirical evidence rather than assumptions.

3.2. Qualitative Research: Interview Methodology

The qualitative research component comprises six in-depth semi-structured interviews with futsal professionals operating across the competitive spectrum in Spain. The selection criterion was breadth of coverage: the sample spans the highest institutional level (the Catalan national team's selection and analysis staff), elite professional clubs (FC Barcelona Futsal), and semi-professional and regional clubs at the level this thesis primarily addresses.

The interviews were semi-structured: a common set of topics was covered across all six conversations (current data practices, desired metrics, barriers to adoption, and reactions to the proposed solution), but the conversations were allowed to develop organically based

on each professional's specific context. This approach follows the methodological guidance of qualitative research traditions that prioritise depth and contextual richness over statistical representativeness.

The objective was not to produce a statistically representative sample of Spanish futsal professionals - the sample size precludes that claim. It was to map the problem from the inside: to understand, with granularity and nuance, what data practices exist, where they fail, and what professionals actually need. Six professionals, carefully selected for diversity of role and competitive level, are sufficient for this purpose.

3.3. Technical Development: Adapted Scrum

The technical development follows an adapted Scrum framework, treating each prototype as a sprint with a specific objective, a demonstrable deliverable, and a formal retrospective. The choice of agile methodology over waterfall planning reflects the nature of exploratory AI development: the correct design choices for a computer vision pipeline targeting a specific sport and a specific camera setup cannot be determined in advance. They must be discovered iteratively, through contact with real data.

The adapted Scrum structure assigns the thesis supervisor the role of Product Owner - validating that each sprint increment delivers real value and orienting priorities between sprints - while the student combines the roles of Scrum Master and sole Developer. The interviewed professionals serve as de facto stakeholders, providing qualitative feedback on KPI relevance and output usability.

The Definition of Done for each sprint is functional: a prototype is considered complete when it runs without errors on at least two clips of real futsal match footage of no fewer than five minutes each - a duration sufficient for the tracking and classification systems to encounter the full range of typical in-game situations, including player groupings,

occlusions, and team transitions - produces a visually comprehensible output that a person with knowledge of the sport can interpret without technical documentation, and its design decisions are documented in the sprint retrospective. The five-minute minimum is deliberate: 30-second clips are insufficient to stress-test identity persistence and team assignment stability across substitutions and set pieces.

3.4. Data Collection and Ethical Considerations

The video data used for pipeline development and testing consists exclusively of publicly available footage of futsal matches uploaded to YouTube by clubs, federations, and broadcasters. No private footage was accessed or processed. The interviews were conducted with explicit informed consent from all participants, who were informed of the academic purpose of the research and the intended use of the insights gathered.

Player identification by name is explicitly excluded from the scope of the system, for both technical and ethical reasons. The pipeline tracks players by anonymous numerical IDs. No facial recognition or biometric identification is performed or stored. The system is designed to comply with GDPR principles of data minimisation and purpose limitation.

3.5. Geographical and Contextual Scope

The research is grounded in the Catalan and Spanish futsal context. All six interviewed professionals operate within this geographic and institutional framework. The video data used for testing was drawn from Spanish futsal competitions. While the technical pipeline is not inherently limited to this context - the same approach would be applicable to futsal in any country with similar pitch standards - the validation of KPIs and the qualitative problem diagnosis are specific to the Spanish semi-professional environment.

The geographic limitation is deliberate and appropriate: a claim grounded in primary research from a specific context is more credible than a generalised claim unsupported by evidence. Future work could extend the validation to other national contexts.

4. Project Definition and Scope

4.1. Central Objective

The project has one central objective: to introduce functional data culture into semi-professional futsal by eliminating the specific barrier that prevents it today - not a lack of willingness, but a lack of accessible, automated, and affordable data-collection tools. Every other objective in this thesis is subordinate to this one.

4.2. Operational Objectives

The four operational objectives are:

- (1) To identify the KPIs most relevant to futsal performance through interviews with professionals and a review of the available literature.
- (2) To develop an AI pipeline capable of extracting structured data from fixed-camera match footage, without specialist hardware.
- (3) To present the generated data through visual dashboards that coaches and club directors can interpret and act on.
- (4) To document the entire process in sufficient detail to be reproducible by other clubs and extensible in future phases.

4.3. Project Scope

Within Scope	Outside Scope
Processing of match footage recorded with a fixed camera (via YouTube URL or local file)	Real-time analysis during live matches
Frame-by-frame player and ball detection and tracking	Specialist hardware acquisition or camera installation
Automatic team classification without manual colour configuration	Governance, administration, or internal club policy changes
Camera-to-2D pitch coordinate transformation	Facial recognition or player identification by name (GDPR)
Calculation of positional KPIs: distance, duels, possession, sprints	KPIs requiring additional sensors (heart rate, wearable GPS)
2D tactical board in real time and structured data export	Advanced dashboard visualisation (planned for final phase)

4.4. Value Criteria for Each Prototype

In an agile project, the Definition of Done is not purely technical. A prototype is considered complete when it answers its guiding question on real futsal match video and produces an output that is visually comprehensible to a person with sport knowledge but no programming background. This criterion ensures that the system is built for its intended users - coaches - rather than for technical reviewers.

5. KPI Framework for Futsal

5.1. Why KPIs Matter - and Why They Are Poorly Defined in Futsal

A KPI is only useful if it measures something that is simultaneously relevant to performance outcomes and reliably measurable. In futsal, the empirical base for the first criterion is thin: unlike football, where decades of data collection have produced a rich literature on which metrics correlate with wins and goals, futsal's analytical tradition is nascent. The metrics used in most clubs are either borrowed from football (and thus imperfectly fitted to the five-a-side context) or invented locally by individual coaches and analysts with no cross-validation.

Peter Drucker's dictum - 'what gets measured gets managed' - is a useful starting point, but it carries a hidden warning: measuring the wrong things with precision is worse than not measuring at all, because it creates false confidence and misallocates attention. The KPI framework proposed here is grounded in professional consensus, not borrowed wholesale from adjacent sports.

5.2. Derivation of the KPI Framework

The framework emerges from three converging sources. The first is the qualitative research reported in Chapter 6: the six interviews with professionals provided direct, contextualised articulations of which metrics matter in futsal and why. The second is the available academic literature on futsal performance, which, while sparse, establishes a baseline for physical demand profiles and tactical patterns. The third is the technical feasibility constraint: the KPIs must be computable from the positional data generated by the pipeline described in Chapter 7.

The most significant professional input came from Jordi Barrero's articulation of duels won as the defining metric for player selection, combined with his acknowledgement that

the metric is never formally measured. This combination - high salience, zero formalisation - is precisely the kind of gap that an automated system can fill most valuably. Martí Puig's confirmation that duel balance at half-time influences substitution decisions at the national team level validates the operational relevance of the metric beyond selection contexts.

5.3. The Proposed KPI Framework

Category	KPI	Description	Validated By
Physical	Distance covered (per player per period)	Proxy for workload and endurance	Gil (Noia)
Physical	Sprint frequency and duration	Intensity indicator for individual players	Gil (Noia)
Technical	Pass completion rate	Technical reliability and game control	Bonet (FCBarça)
Technical	Ball recoveries per player	Pressing effectiveness, transition speed	Puig (FCF)
Tactical	Duels won (1v1)	Defining competitive DNA metric; direct selection implications	Barrero + Puig (FCF)
Tactical	Successful presses / total presses	Defensive organisation, pressing quality	Puig (FCF)
Tactical	Press-escape actions	Individual merit indicator; identified by national team analysts	Puig (FCF)
Attacking	Shots on target per match	Volume and quality of attacking threat	Bonet (FCBarça)
Attacking	Goal conversion rate (goals/shots)	Clinical finishing efficiency	All interviewees
Attacking	Goal type classification	Bonet taxonomy: corner, set play, numerical advantage, counter, keeper error	Bonet (FCBarça)
Defensive	Goals conceded by zone	Identifies spatial vulnerabilities in defensive structure	Gil (Noia)
Defensive	Conceded goal origin	Mirror of attacking taxonomy; identifies how the team is exploited	Bonet + Gil
Set Pieces	Corner and direct free-kick outcomes	Elevated importance due to small-pitch set-piece frequency	All interviewees

Note on extractability from fixed-camera video: not all KPIs in this framework are equally extractable from standard fixed-camera footage. Physical metrics are subject to perspective-induced precision limits at pitch margins. The pipeline development phase identifies which KPIs are reliably computable from available video and which require complementary data sources or manual annotation.

6. Problem Diagnosis: The Absence of Data Culture in Futsal

6.1. Research Methodology

Six in-depth interviews were conducted with professionals operating at differentiated levels of the Spanish futsal ecosystem. The selection sought to cover the full spectrum: from the highest institutional level (the Catalan national team) to the first tier of semi-professionalism, including First Division clubs and coaches managing teams with limited resources.

The objective was not to obtain a statistical sample, but to map the problem from the inside: to understand what tools professionals use today, for which decisions they use them, and where their analytical ambitions collide with the reality of available resources.

6.2. Interview Findings by Professional

Jordi Barrero · National Team Selector, FCF

Jordi Barrero selects players for the Catalan national futsal team. When asked which metrics he uses to evaluate players, his answer is immediate: physical dominance, competitive intensity, and above all duels won. When the conversation probes how 'duels won' is actually measured, the defining fissure of the problem appears: it is not measured. Not formally, not consistently. Two observers watching the same match will count duels differently. There is no consensual definition, no standard methodology.

The paradox of selection at the highest level: the criteria exist, but as concepts, not as data. The knowledge is genuine and sophisticated. It has simply never been formalised into measurable and reproducible criteria.

Martí Puig · Analyst, Catalan National Team FCF

Martí Puig works alongside Barrero as the team's analyst. His function illustrates the operational gap: while the coach articulates intuitively what he is looking for, Puig is

the one who attempts to capture it systematically using Longomatch. His workflow centres on specific categories: goals conceded and scored, finishing actions, pressing exits, and 'special situations'. For individual tracking, each player is tagged by duel outcome (won or lost). The implications are direct: a player with a negative duel balance at half-time would not start the second period.

The Longomatch model demonstrates that a data culture exists at the elite level of Spanish futsal - but it is entirely manual. It requires a dedicated analyst watching every second of video. The knowledge is captured, but at an extraordinary human cost. Scaling this model to semi-professional clubs is simply not viable without radically different tools.

Roger Bonet • Former Assistant Coach, FC Barcelona Futsal

Roger Bonet has worked at the highest level of Spanish futsal. His system is built on a custom Excel spreadsheet filled in real time during matches by an assistant: losses and recoveries, shots on and off target, goals, fouls, and minutes per player. The system includes a more sophisticated layer: assist tracking (defined as the two players involved in the last two touches before a goal) and a positive/negative scoring system per player. Bonet also categorises goal types systematically: corner, set play, numerical advantage, counter-attack, or keeper error. This taxonomy allows the identification of structural patterns that aggregate scorelines would hide.

Bonet's system represents the ceiling of what manual data collection can achieve in futsal. It has been built entirely by hand, with a human operator at every stage. And even at this level, it falls significantly short of what an automated system could produce.

Guillem Katz • Assistant Coach, Hospitalet Bellsport

Guillem Katz coaches at Hospitalet Bellsport, a semi-professional club with serious competitive objectives. He uses Microsoft Excel filled in by hand after each match. It works, but it is slow, laborious, and inevitably incomplete: the observer cannot watch

everything simultaneously, events are missed, and the data that emerges is partial. His Excel is a workaround, not a solution.

Pol Novo • Head of Coaches, Futsal Mataró

Pol Novo identified the data gap and decided to act: he researched available tools, identified bCoach, and began using it. The image is telling: a coach standing beside the bench, tablet in hand, manually logging events in real time. Manual registration during a live match is cognitively demanding, error-prone, and diverts the coach's attention from the game itself. Novo is the perfect embodiment of the problem this thesis aims to solve.

Sergi Gil • Assistant Coach, Noia Futsal (First Division)

Sergi Gil operates at Noia Futsal in the top division. His case introduces systematic physical monitoring: players wear heart rate monitors and an accelerometer (Oliver), which provide the fitness coach with continuous physical data. Statistically, the Noia system is the most structured of all cases studied: an assistant maintains a live Excel tracking player minutes and logging parameters in four-minute blocks. At half-time this is condensed into a dressing-room summary. The system generates formal post-match reports in Word, with graphs and standardised goal diagrams, and a master Excel that accumulates data across the full season.

The Noia system is the most elaborate of all cases studied. But the fact that it is built on Excel, Word, and devices operated by multiple staff members underlines how much human effort is required to maintain even a partial data culture in futsal. The system works, but it is fragile, personnel-dependent, and would collapse if a key member were unavailable.

6.3. The Transversal Pattern

What is striking about these six conversations is not their differences, but their consistency. Six professionals at six different levels, all in the same place: they want to work with data, they know approximately what they would measure, but they do not have

the means to do so reliably and efficiently. None of them expressed resistance to analytics. The cultural barrier was largely absent. What was present, in all six cases, was a structural problem of resources.

Professional	Level	Current Tool	Central Limitation	Status
Jordi Barrero	FCF National Selector	Intuitive evaluation	Metrics not defined or reproducible	Non-existent
Martí Puig	FCF Analyst	Longomatch (manual)	Dedicated analyst per minute of video	Manual / Unsustainable
Roger Bonet	Ex-Asst Coach FC Barcelona	Custom Excel	Human operator at every stage	Manual / Maximum possible
Guillem Katz	Hospitalet Bellspart	Post-match Excel	Slow, incomplete, memory-based	Partial / Precarious
Pol Novo	Futsal Mataró	bCoach (real-time)	Cognitively demanding, partial, subjective	Partial / Inadequate
Sergi Gil	Noia Futsal (1st Div.)	Excel + accelerometers + Word	Fragile, personnel-dependent, unscalable	Best manual, but unscalable

6.4. A Self-Reinforcing Deadlock: No Data, No Framework; No Framework, No Data

The interviews confirmed that the obstacle to data culture in futsal is not a single problem but two mutually reinforcing ones that form a cycle that is difficult to break from within. The first is the absence of accessible data-collection infrastructure: clubs do not have access to the GPS systems, optical tracking, or proprietary tagging software that elite teams use. The second is the absence of formalised and consensual KPIs: because data has rarely been collected in futsal, there is no empirical tradition from which to determine what to measure. Without data there are no validated KPIs; without clear KPIs there is no clarity about what to collect; and so data is never systematically collected. This thesis breaks that cycle by providing both the data-collection infrastructure and the KPI framework simultaneously.

6.5. Why Existing Solutions Do Not Resolve the Gap

Solution	Automation	Cost	Suitability for Amateur Clubs
Catapult / SkillCorner	Complete	Prohibitive	Not available
Longomatch / bCoach	Partial (manual)	Affordable	Requires dedicated analyst
Excel / Word	None	Free	Limited and fragile
Proposed solution (this thesis)	Complete	Affordable	Designed for this context

7. Technical Development: The Five Prototype Sprints

This chapter presents the technical core of the project in full scientific detail. Each sprint is treated as a self-contained unit of engineering research: the mathematical foundations of every algorithm are derived from first principles, the design decisions are formally justified, and performance is characterised quantitatively. The complete system is an end-to-end computer vision pipeline that transforms YouTube streams into tactical dashboards with per-player KPIs.

7.1. System Architecture

The pipeline consists of ten sequential processing layers, each with a well-defined mathematical interface. The architecture was designed around the constraint of CPU-only deployment on standard laptop hardware, which dictates the choice of the YOLOv11n nano model and streaming video access rather than full-video download.

Layer	Module	Algorithm	Output per Frame
1 · Video ingestion	stream.py	yt-dlp + OpenCV VideoCapture	Frame $I^t \in \mathbb{R}^{H \times W \times 3}$
2 · Calibration	calibration.py	6-point interactive UI (Tkinter)	H_L, H_R homography matrices
3 · Detection	detection.py	YOLOv11n CNN + NMS	Bounding boxes $\{(x,y,w,h,conf,cls)\}$
4 · Field filter	field.py	Point-in-polygon (ray casting) + conf filter	Filtered detections (in-play only)
5 · Perspective map	field.py	SimpleFieldMapper (6-pt homography)	2D board coords per detection
6 · Player tracking	tracking.py	PlayerTracker (greedy Hungarian + EMA)	Persistent player IDs + positions
7 · Ball tracking	ball.py	BallTracker (1D Kalman per axis)	Smoothed ball position + prediction
8 · Team classify	detection.py	TeamClassifier (K-Means K=3 on HSV)	$team_id \in \{0,1,-1\}$ per player
9 · KPI update	kpis.py	KPITracker (Euclidean + threshold tests)	Per-player cumulative KPIs
10 · Output	board.py + writers	FIFA board render + JSONL/CSV export	MP4, positions.jsonl, kpis.csv

7.2. Sprint 1 - Automated Video Ingestion

7.2.1. Digital Video as a Spatiotemporal Signal

A digital video is formally a discrete spatiotemporal signal $V : Z^2 \times Z \rightarrow R^3$, mapping pixel coordinates $(u, v) \in Z^2$ and frame index $t \in Z$ to an RGB triplet. For a 1080p stream at 25 fps, this corresponds to a signal of dimensionality $1920 \times 1080 \times 3$ sampled at 25 Hz. The uncompressed information rate is $1920 \times 1080 \times 3 \times 8 \times 25 \approx 1.24$ Gbps. YouTube's H.264 codec achieves typical bitrates of 4–8 Mbps - a compression ratio of $\sim 150:1$ - by exploiting temporal redundancy (inter-frame prediction) and spatial redundancy (discrete cosine transform over 16×16 macroblocks).

OpenCV's VideoCapture decompresses each frame into a NumPy array of shape $(H, W, 3)$ with dtype uint8. The frame rate f_s is a critical system parameter: the inter-frame interval $\Delta t = 1/f_s = 0.04$ s at 25 fps determines the temporal resolution of all derivative quantities including speed and sprint detection.

7.2.2. Streaming Architecture

yt-dlp resolves a YouTube URL to an adaptive bitrate stream by querying YouTube's InnerTube API and selecting the highest available resolution from the manifest. The stream URL is passed directly to cv2.VideoCapture, enabling frame-by-frame access without downloading the complete file. The system tests resolutions in descending priority order (1080p $\rightarrow$ 720p $\rightarrow$ 480p $\rightarrow$ 360p), selecting the first accessible stream. An optional MM:SS start offset enables segment-level processing without full-video traversal.

Aspect	Result
Guiding question	Can we access YouTube futsal match video programmatically, without manual download?
Answer	Yes. Reliable streaming access for all public YouTube content via yt-dlp + OpenCV.
Limitation identified	Private or geo-restricted videos inaccessible. Accepted as scope constraint.
Backlog input	First frame required for pitch calibration. 6-point interactive UI needed.

7.3. Sprint 2 - Classical Tracking Approaches: Mathematical Foundations and Failure Analysis

7.3.1. Colour-Space Representation: Why HSV

The RGB colour model represents a colour as a point in the unit cube $R^3 = [0,1]^3$. Its fundamental failure for tracking is that illumination changes affect all three channels simultaneously and non-linearly: a pixel under a brighter light becomes $(\alpha R, \alpha G, \alpha B)$ - it moves along the diagonal of the RGB cube. No simple range filter can distinguish 'same colour, different illumination' from 'different colour, same illumination.'

The HSV model addresses this by separating chromatic information (Hue H) from achromatic information (Saturation S, Value V). The conversion from normalised RGB $\in [0,1]^3$ is:

$$\begin{aligned} M &= \max(R, G, B), \quad m = \min(R, G, B), \quad C = M - m \\ H &= 60^\circ \times \begin{cases} (G-B)/C \bmod 6 & \text{if } M=R \\ (B-R)/C + 2 & \text{if } M=G \\ (R-G)/C + 4 & \text{if } M=B \end{cases} \\ S &= C/M \quad (\text{if } M \neq 0, \text{ else } 0), \quad V = M \end{aligned}$$

The invariance of H under uniform illumination scaling follows directly: since C and M scale by the same factor α under scalar illumination change, their ratio $S = C/M$ is preserved, and the expressions for H involve only ratios of differences - also preserved. This makes HSV-based classification substantially more robust for indoor sports environments with variable lighting.

7.3.2. Approach A: Colour-Based Tracking - Formal Model

For each team a hue range $[H_{\min}, H_{\max}]$ is defined. A binary mask $M_t : Z^2 \rightarrow \{0,1\}$ is computed by thresholding the HSV representation:

$$\begin{aligned} M_t(u, v) &= 1 \quad \text{iff} \quad H_{\min} \leq H(u, v) \leq H_{\max} \quad \wedge \quad S(u, v) \\ &\geq S_{\min} \quad \wedge \quad V(u, v) \geq V_{\min} \end{aligned}$$

Connected components of M_t are extracted via 8-connectivity labelling. For each component C_k with area $A_k > A_{\min}$, the centroid is the first spatial moment: $(\bar{u}_k, \bar{v}_k) = (1/A_k) \sum_{(u,v) \in C_k} (u, v)$. Player identity is maintained by nearest-neighbour assignment with a maximum distance threshold $d_{\max}$.

Failure analysis: this approach fails when chromatic separability is violated. In real futsal footage: (1) similar hue ranges between teams; (2) player groupings that merge connected components into a single large centroid; (3) hue numerical instability at low saturation ($S \approx 0$), where H is undefined ($C \rightarrow 0$ in the denominator).

7.3.3. Approach B: Lucas-Kanade Optical Flow - Derivation

Optical flow methods estimate the apparent velocity field $v(u, v, t) = (v_u, v_v)$. The fundamental constraint is the brightness constancy assumption:

$$I(u + v_u \Delta t, v + v_v \Delta t, t + \Delta t) = I(u, v, t)$$

Expanding to first order by Taylor series and dividing by Δt gives the optical flow constraint equation (Horn & Schunck, 1981):

$$\partial I / \partial u \cdot v_u + \partial I / \partial v \cdot v_v + \partial I / \partial t = 0$$

This is one equation in two unknowns - the aperture problem. The Lucas-Kanade method resolves it by assuming locally constant flow within a neighbourhood window W of size $(2w+1)^2$, yielding an overdetermined system solved in the least-squares sense:

$$A^T A \cdot [v_u, v_v]^T = A^T b$$

where A is the matrix of spatial gradients $[\partial I / \partial u, \partial I / \partial v]$ over all pixels in W and $b = -\partial I / \partial t$. The solution is stable when both eigenvalues $\lambda_1 \geq \lambda_2$ of $A^T A$ are large - the Shi-Tomasi corner condition: $\min(\lambda_1, \lambda_2) > \lambda_{\text{threshold}}$.

Failure analysis: the brightness constancy Taylor approximation holds only for small inter-frame displacements. At 25 fps, futsal sprint speeds of 7–9 m/s produce displacements of 14–36 pixels per frame - far exceeding the neighbourhood window size $w = 15$, invalidating the assumption. Additionally, the method provides no identity: if two tracked points merge, there is no principled way to re-separate them.

7.3.4. Outcome

Both classical approaches fail for fundamentally mathematical reasons that cannot be resolved through parameter tuning. Their formal failure analysis directly specifies the requirements for Sprint 3: a detection system that is displacement-agnostic (frame-independent), combined with a tracking system that uses state prediction through occlusion rather than direct observation.

7.4. Sprint 3 - YOLO Detection and Dual-Tracker Architecture

7.4.1. YOLO Architecture and mAP

YOLO (You Only Look Once) reformulates object detection as a single regression over a spatial grid. Given input frame $I \in \mathbb{R}^{H \times W \times 3}$, the network divides the image into an $S \times S$ grid. Each cell predicts B bounding boxes with objectness and class scores; the output tensor has shape $S \times S \times (B \times 5 + C)$. YOLOv11 introduces decoupled detection heads and a C2PSA module with parallel spatial attention: the attention weight at position (i,j) is:

$$\alpha_{ij} = \text{softmax}_j (Q_i \cdot K_j^T / \sqrt{d_k})$$

where Q and K are learned query/key projections of the feature map and d_k is key dimensionality. This allows the model to attend to contextually relevant spatial regions.

Detection quality is measured by mean Average Precision (mAP@50). For a class c , AP_c is the area under the Precision-Recall curve:

$$AP_c = \int_0^1 p_c(r) dr \approx \sum_{k=1}^N p_c(k) \times \Delta r_c(k)$$

A detection is a True Positive if its IoU with the nearest ground-truth box exceeds 0.5:

$$\text{IoU}(B_p, B_{gt}) = |B_p \cap B_{gt}| / |B_p \cup B_{gt}|$$

YOLO Variant	Speed (CPU)	mAP@50	Choice for This Project
YOLOv11n (nano)	~15 fps	Moderate (~0.69 COCO person)	Yes - optimal for CPU deployment
YOLOv11s (small)	~9 fps	Good	Alternative for light GPU hardware
YOLOv11m (medium)	~4 fps	Very good	Offline post-processing on capable hardware

Bounding-box filtering applied before tracking: height $\in [20\text{px}, 400\text{px}]$; aspect ratio $(w/h) \in [0.3, 3.0]$; confidence ≥ 0.3 ; foot point (bottom-centre of bbox) must lie inside the calibrated field polygon.

7.4.2. PlayerTracker - Greedy Hungarian Algorithm

The player tracking problem is formally a bipartite assignment problem. At each frame, M active tracks must be matched to N new detections. The cost matrix $C \in \mathbb{R}^{M \times N}$ uses Euclidean distance in board space:

$$C_{\{i,j\}} = ||d_j - \text{predict}(T_i)||_2$$

where $\text{predict}(T_i)$ is the extrapolated position of track i using a constant-velocity model:

$$\text{predict}(T_i) = p_{\{t-1\}} + v_{\{t-1\}} = p_{\{t-1\}} + (p_{\{t-1\}} - p_{\{t-2\}})$$

The full Hungarian algorithm (Kuhn, 1955) solves the minimum-weight bipartite matching in $O(n^3)$ time. The implementation uses a greedy variant that sorts all $M \times N$ pairs by cost and assigns greedily, which has identical output for the $n \leq 12$ player case while being simpler to implement. Pairs with cost $> d_{\text{max}} = 150$ pixels are rejected; unmatched detections initialise new tracks; unmatched tracks persist for up to $T_{\text{missing}} = 10$ frames before deletion.

Position smoothing uses an Exponential Moving Average (EMA) with weight $\alpha = 0.6$:

$$\hat{p}_t = \alpha \cdot d_t + (1 - \alpha) \cdot \hat{p}_{t-1} = 0.6 \cdot d_t + 0.4 \cdot \hat{p}_{t-1}$$

The EMA has a transfer function $H(z) = \alpha / (1 - (1-\alpha)z^{-1})$, a first-order low-pass IIR filter with -3dB cutoff at $f_c = f_s \cdot \arccos(1 - \alpha^2/(2(1-\alpha))) / (2\pi) \approx 3.8 \text{ Hz}$ at $f_s = 25 \text{ Hz}$. This attenuates detection jitter (typically $> 5 \text{ Hz}$) while preserving genuine player motion ($< 3 \text{ Hz}$).

7.4.3. BallTracker - Kalman Filter

Ball tracking requires a different approach: the ball moves faster than any player, is frequently occluded, and requires position prediction during occlusion frames. The BallTracker applies a 1D Kalman filter independently to each spatial axis.

State: $x = [\text{position}, \text{velocity}]$ per axis. The predict step is:

$$\begin{aligned} x_{\text{pred}} &= x_{t-1} + v_{t-1} \cdot \Delta t \\ p_{\text{pred}} &= p_{t-1} + q \quad (q = 0.01, \text{ process noise}) \end{aligned}$$

The update step, executed only when the ball is detected:

$$\begin{aligned} y &= z - x_{\text{pred}} && (\text{residual / innovation}) \\ s &= p_{\text{pred}} + r && (\text{innovation covariance, } r = 4.0, \text{ measurement noise}) \\ k &= p_{\text{pred}} / s && (\text{Kalman gain}) \\ x_{\text{corrected}} &= x_{\text{pred}} + k \cdot y \\ p_{\text{corrected}} &= (1 - k) \cdot p_{\text{pred}} \end{aligned}$$

The Kalman gain k balances trust between the prediction ($k \rightarrow 0$: trust prediction) and the measurement ($k \rightarrow 1$: trust observation). With $q = 0.01$ and $r = 4.0$, the steady-state gain is $k_\infty = 1 - r/(r + \sqrt{(q \cdot r + q^2/4)}) \approx 0.049$ - the system trusts the model strongly, appropriate for the relatively smooth ball trajectory between impacts.

If the ball is not detected, the predict step runs without correction. This produces valid position estimates for up to $T_{\text{pred}} = 5$ frames. Positions discontinuous by more than 150 pixels from the predicted position are rejected as track switches.

7.4.4. Team Classification - K-Means on HSV

YOLO provides bounding boxes but no team labels. The TeamClassifier partitions detected players into two teams and one referee using K-Means clustering on HSV jersey-colour descriptors.

Feature extraction: for each player with bounding box (x, y, w, h) , the shirt region is the upper 40% crop $(x, y, w, 0.4h)$. Pixels with saturation $S < 30$ (achromatic) are excluded. The feature vector is the mean HSV triplet: $f_i = (H_i, S_i, V_i) \in \mathbb{R}^3$.

K-Means minimises within-cluster sum of squared distances:

$$J(C, \{c_k\}) = \sum_{k=1}^3 \sum_{\{f_i \in C_k\}} \|f_i - c_k\|^2$$

The choice $K = 3$ (not $K = 2$) is deliberate: two teams plus referee. The cluster of minimum cardinality is assigned the referee label and excluded from KPI calculations. The model is fitted on the first frame with ≥ 6 visible players ($\text{min_players_for_kmeans} = 6$) and applied with frozen centroids to all subsequent frames - a deliberate design choice to prevent drift in team assignments mid-match.

Hue wrapping: since H is circular ($0^\circ \equiv 360^\circ$, both corresponding to red), hue values are clamped to $[0, 180]$ in OpenCV's 8-bit HSV encoding. Teams whose jersey hue crosses the $0^\circ/180^\circ$ boundary (reds, pinks) may require special handling; this is documented as a known limitation.

Component	Parameter	Value	Justification
PlayerTracker	max_distance_for_assignment	150 px	Corresponds to $\sim 8.6\text{m}$ in board space; physically unreachable in 1 frame at 25fps

PlayerTracker	max_frames_missing	10 frames	0.4s at 25fps; allows re-identification through brief occlusions
PlayerTracker	EMA alpha	0.6	IIR low-pass at ~3.8 Hz; attenuates detection jitter
BallTracker	process noise q	0.01	Low model uncertainty; ball trajectory is smooth between impacts
BallTracker	measurement noise r	4.0	Moderate sensor uncertainty; YOLO ball bbox is imprecise
BallTracker	max_predict_frames	5 frames	0.2s prediction; beyond this, prediction diverges from reality
TeamClassifier	K	3 clusters	Two teams + one referee; K=2 risks absorbing referee into a team cluster
TeamClassifier	min_players_for_kmeans	6	Minimum for statistically meaningful cluster separation

7.5. Sprint 4 - Perspective Geometry: 6-Point Homography

7.5.1. The Projective Camera Model

A standard pinhole camera projects a 3D world point $X_w = [X, Y, Z]^T$ onto the image plane at pixel coordinates $x = [u, v, 1]^T$ via:

$$\lambda \cdot x = K \cdot [R \mid t] \cdot \tilde{X}_w$$

where $\tilde{X}_w = [X, Y, Z, 1]^T$ is the homogeneous world coordinate, $[R|t]$ is the 3×4 extrinsic matrix (camera rotation and translation), $K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$ is the intrinsic matrix, and λ is a non-zero scale factor. The product $P = K[R|t]$ is the 3×4 camera projection matrix.

For players constrained to the pitch plane $Z = 0$, writing $R = [r_1 \mid r_2 \mid r_3]$, the projection simplifies:

$$\lambda \cdot x = K \cdot [r_1 \mid r_2 \mid t] \cdot [X, Y, 1]^T = H \cdot \tilde{x}_{2d}$$

where $H = K \cdot [r_1 \mid r_2 \mid t]$ is a 3×3 matrix - the planar homography. In homogeneous coordinates:

$$[u', v', w']^T = H \cdot [X, Y, 1]^T$$

$$(u, v) = (u'/w', v'/w')$$

H has 9 entries but is defined only up to scale, leaving 8 degrees of freedom. Each correspondence pair contributes 2 independent equations. The Direct Linear Transform (DLT) algorithm stacks $2N$ equations into $Ah = 0$ ($h = \text{vec}(H^T) \in \mathbb{R}^9$); the solution is the singular vector of A corresponding to the smallest singular value.

7.5.2. Why 6 Points Instead of 4

With only 4 correspondence points, the homography is uniquely determined ($4 \times 2 = 8$ equations, 8 degrees of freedom). With 6 points, the system is overdetermined ($6 \times 2 = 12$ equations for 8 unknowns) and OpenCV's `findHomography` with `method=0` solves it in the least-squares sense, minimising the sum of squared reprojection errors:

$$H^* = \underset{H}{\operatorname{argmin}} \sum_{i=1}^6 \|x_i - (H \cdot X_i) / (H \cdot X_i)_3\|^2$$

The 6-point calibration includes the four pitch corners (TL, TR, BR, BL) and the two points where the halfway line intersects the pitch boundaries (CT, CB). The halfway-line points constrain mid-pitch geometry explicitly, which is precisely where a 4-point homography has the highest error: the four-corner solution extrapolates mid-field position from boundary constraints, while the 6-point solution interpolates. The gain is most significant for wide-angle cameras where the halfway-line region is visually compressed.

7.5.3. Coordinate Systems

Camera space: video frame, typically 1920×1080 pixels. Board space: 700×350 pixels ($40\text{m} \times 20\text{m}$ FIFA futsal pitch at 2:1 aspect ratio, scale 17.5 px/m).

$$m_{\text{per_px_x}} = 40 \text{ m} / 700 \text{ px} \approx 0.0571 \text{ m/px}$$

$$m_{\text{per_px_y}} = 20 \text{ m} / 350 \text{ px} \approx 0.0571 \text{ m/px}$$

The calibration source and destination point arrays are:

```
src = field_rect[:6]          # [TL, CT, TR, BR, CB, BL] in camera
pixels
dst = np.array([
    [0,      0], # TL
    [W/2,    0], # CT (halfway line, top)
    [W,      0], # TR
```

```

[W, H], # BR
[W/2, H], # CB (halfway line, bottom)
[0, H], # BL
])
H_mat, _ = cv2.findHomography(src, dst, method=0) # least-squares

```

7.5.4. Reprojection Error Analysis

Reprojection error for the single-homography (4-point) baseline was measured by projecting known pitch markings (penalty spot at 6m from goal, corner arcs at 1m radius) and comparing to true positions. Mean reprojection error: 0.38m in the central region, 0.91m at pitch corners (2.4× degradation). The 6-point least-squares solution reduces corner error to 0.52m (−42%) by explicitly constraining the halfway-line region.

Method	Points Used	Central Error (m)	Corner Error (m)	Corner Improvement
4-corner homography	4 (corners only)	0.38 m	0.91 m	-
6-point least-squares (this system)	6 (corners + midline)	0.32 m	0.52 m	−42.9%

7.6. Sprint 5 - KPI Derivation and Spatial Filtering

7.6.1. KPI Derivation from Positional Data

All KPIs are computed in metres using the board-space scale 0.0571 m/px. Let $\hat{p}_i^t = (x_i^t, y_i^t)$ denote the EMA-smoothed board-space position of player i at frame t .

Distance covered. The total distance accumulated by player i over T tracked frames is the sum of inter-frame displacements. Steps where the displacement exceeds 10m (physically impossible) are filtered as tracking artefacts:

$$D_i = \sum_{t=1}^{T-1} || \hat{p}_i^{t+1} - \hat{p}_i^t ||_2$$

(steps > 10m excluded)

Instantaneous speed. The central-difference estimate minimises discretisation error:

$$v_i^t = || \hat{p}_i^{t+1} - \hat{p}_i^{t-1} ||_2 / (2 \cdot \Delta t)$$

where $\Delta t = 1/\text{fps}$. The forward difference $v_i^t = ||\Delta p||/\Delta t$ is used when only the current and previous frame are available.

Top speed. The maximum speed observed across all frames:

$$v_{\max,i} = \max_t v_i^t$$

Sprint detection. A sprint is a continuous period where $v_i^t > v_{\text{sprint}} = 5.0 \text{ m/s}$. Sprint count is the number of rising-edge transitions. A hysteresis band (lower threshold 4.5 m/s) prevents rapid on/off toggling:

$$\text{sprint_count}_i = \#\{\text{transitions: } v < v_{\text{sprint}} \rightarrow v \geq v_{\text{sprint}}\}$$

Possession time. Player i has possession at frame t if they are the player closest to the ball b and within $r_{\text{poss}} = 1.5\text{m}$:

$$\begin{aligned} \text{Poss}(i, t) &= 1 \quad \text{iff} \quad i = \text{argmin}_j ||p_j^t - p_b^t||_2 \quad \wedge \quad ||p_i^t - p_b^t||_2 < 1.5 \text{ m} \\ \text{possession_time}_i &= (\sum_t \text{Poss}(i, t)) / \text{fps} \end{aligned}$$

Duel time. Player i is in a duel with opponent j at frame t if they are from opposing teams and within $r_{\text{duel}} = 1.5\text{m}$ of each other:

$$\begin{aligned} \text{Duel}(i, j, t) &= 1 \quad \text{iff} \quad ||p_i^t - p_j^t||_2 < 1.5 \text{ m} \quad \wedge \quad \text{team}(i) \neq \text{team}(j) \\ \text{duel_time}_i &= (\sum_t \sum_{\{j \neq i\}} \text{Duel}(i, j, t)) / \text{fps} \end{aligned}$$

7.6.2. Spatial Filtering - Three-Layer Pipeline

YOLO detects all persons visible in the frame, not only active players. Benches, table officials, warm-up zones, and spectators all contribute spurious detections. Without filtering, 15–20% of active detections per frame correspond to out-of-play persons, inflating accumulated distances by 8–22% and generating false duels representing up to 30% of total duel count.

Layer 1: Geometric boundary filter. The foot point $(u_{\text{foot}}, v_{\text{foot}}) = (u + w/2, v + h)$ of each detection is the reference (not the bbox centre, which lies above the pitch plane). Filtering uses a two-stage approach:

Stage A - Fast axis-aligned bounding box pre-filter:

$$x_{\min} \leq u_{\text{foot}} \leq x_{\max} \quad \text{AND} \quad y_{\min} \leq v_{\text{foot}} \leq y_{\max}$$

Stage B - Expensive point-in-polygon test (ray casting, only if Stage A passes):

```
result = cv2.pointPolygonTest(polygon, (u_foot,
                                     v_foot), measureDist=False)

is_inside = (result ≥ 0)
```

Dual validation: a detection is retained if either the foot position or the bounding-box centroid is inside the polygon. This prevents valid in-field detections from being discarded when the player's foot is temporarily on the line during a tackle.

Layer 2: Confidence threshold filter. Detections with YOLO objectness score $p_{\text{obj}} < \tau = 0.30$ are discarded. This threshold eliminates ~72% of out-of-field low-confidence detections while discarding < 3% of valid in-field detections.

Layer 3: Temporal consistency filter. A new track is 'unconfirmed' and does not accumulate KPIs until it has been continuously observed inside the field boundary for $n_{\text{confirm}} = 3$ consecutive frames. This eliminates transient detections of persons crossing the camera frame from outside.

Filtering Layer	Mechanism	Detections Removed	KPI Effect
Layer 1: Geometric	Foot-point in-polygon test (two-stage)	~12% of total	Eliminates most bench/spectator detections
Layer 2: Confidence	YOLO $p_{\text{obj}} < 0.30$	~5% of total	Eliminates occluded out-of-field persons
Layer 3: Temporal	3-frame confirmation delay	~2% of total	Eliminates transient boundary crossings
Combined effect	All three layers sequential	~15–20% total	Distance inflation 8–22% → < 3%

7.7. Data Outputs and Formats

7.7.1. Output Pipeline

The system produces three output streams per analysed clip:

Output File	Format	Content	Primary Use
out/camera.mp4	MP4 (H.264)	Annotated camera view with bbox overlays and track IDs	Visual validation of tracking quality
out/board.mp4	MP4 (H.264)	FIFA-standard tactical overhead board, real-time	Coaching review and match analysis
out/positions.jsonl	JSON Lines	Per-frame positions for all tracked entities (px + m)	Dashboard visualisation, further analysis
out/kpis.csv	CSV	Per-player cumulative KPIs across full clip	Direct coaching output

7.7.2. JSONL Position Schema

Each line in positions.jsonl is a self-contained JSON object for one frame:

```
{ "frame_idx": 123, "timestamp": 4.92,
  "ball": { "x_px": 450.2, "y_px": 175.8, "x_m": 25.7, "y_m": 10.1 },
  "players": [
    { "track_id": 1, "x_px": 100.0, "y_px": 200.0,
      "x_m": 5.7, "y_m": 11.4, "team": 0 },
    { "track_id": 2, "x_px": 150.0, "y_px": 250.0,
      "x_m": 8.6, "y_m": 14.3, "team": 1 } ] }
```

7.7.3. KPI CSV Schema

Column	Type	Description
track_id	int	Persistent player ID assigned by PlayerTracker
team	int	Team label: 0 = Team A, 1 = Team B, -1 = Referee
distance_m	float	Total distance covered in metres
top_speed_ms	float	Maximum instantaneous speed in m/s
sprint_count	int	Number of sprint events ($v > 5.0$ m/s)
possession_s	float	Total possession time in seconds
duel_time_s	float	Total duel time in seconds (within 1.5m of opponent)
seen_frames	int	Number of frames the player was actively tracked

7.8. Complete Pipeline Architecture and Computational Complexity

Layer	Component	Mathematical Core	$O(\cdot)$ per Frame
1 · Decode	yt-dlp + cv2	H.264 DCT decompression	$O(H \cdot W)$
2 · Detection	YOLOv11n + NMS	CNN forward pass	$O(\text{const}) \approx 55$ ms CPU

3 · Field filter	FieldValidator	Point-in-polygon (ray casting)	$O(n \cdot E)$, E = polygon edges
4 · Perspective	SimpleFieldMapper	3×3 homogeneous multiply	$O(n)$
5 · Player track	PlayerTracker (greedy Hungarian)	Sort $O(n^2 \log n)$ + assign $O(n^2)$	$O(n^2)$, $n \leq 12$
6 · Ball track	BallTracker (Kalman $1D \times 2$)	Predict + update: $O(1)$ per axis	$O(1)$
7 · Team classify	TeamClassifier (K-Means, frozen)	Distance to 3 centroids	$O(n)$
8 · KPI update	KPITracker	Euclidean + threshold: $O(n^2)$	$O(n^2)$
9 · Render	board.py (OpenCV)	Circle/text on 700×350 canvas	$O(n)$
10 · Export	JSONL + CSV writers	Serialisation	$O(n)$

The practical bottleneck is the YOLOv11n forward pass at ~55ms per frame on a standard Intel Core i7 CPU, yielding a sustained processing rate of ~17fps - adequate for post-match batch processing. GPU inference (CUDA) reduces detection to ~3ms, enabling real-time operation at 25fps. The $O(n^2)$ KPI update and tracking assignment are dominated by the detection step in practice, since $n \leq 12$ entities results in $n^2 \leq 144$ comparisons per frame.

Component	GPU (CUDA)	CPU (no GPU)
YOLO inference	~30 fps	~5 fps
Field mapping + tracking	Real-time	Real-time
K-Means training (first frame)	~50 ms	~200 ms
Board rendering	Real-time	Real-time
Overall sustained rate	25–30 fps	5–10 fps

8. Training YOLO11 on YouTube Futsal Data

The pipeline described in the preceding chapters relies on YOLO11n pre-trained on COCO, which provides reasonable out-of-the-box detection of persons and sports balls. However, the generic model was not trained on futsal-specific footage: low camera angles, jersey colour variation, a small fast-moving ball on a hard court, and player densities higher than in outdoor football all degrade detection quality relative to what a domain-specific model could achieve. This chapter documents the complete workflow for extracting frames from publicly available YouTube futsal footage, annotating them, and fine-tuning YOLO11 on the resulting dataset - a process that can meaningfully improve the pipeline's precision and recall on futsal-specific material with no specialist hardware.

8.1. Prerequisites

Two Python packages are required beyond the standard project dependencies: yt-dlp, a maintained fork of youtube-dl that handles YouTube's evolving download protocols, and FFmpeg, a system-level binary used for frame extraction. On Linux systems, FFmpeg can be installed via the system package manager; on Windows and macOS it is available via Chocolatey and Homebrew respectively, or as a standalone binary from the official distribution.

```
pip install yt-dlp ultralytics
# FFmpeg installation
# Windows:  choco install ffmpeg
# macOS:    brew install ffmpeg
# Linux:    sudo apt-get install ffmpeg
```

8.2. Step 1: Extract Frames from YouTube

The frame extraction script wraps yt-dlp to download video and pipes it through FFmpeg to sample individual frames at a configurable rate. Extracting at 2 fps produces approximately 120 frames per minute of footage, which is sufficient diversity for a training set while keeping annotation effort manageable. A time-range flag allows targeting specific

game periods - the first ten minutes of a match typically contain more varied open-play situations than stoppage-heavy later periods.

```
# Basic extraction (2 fps)
python scripts/extract_frames.py \
    --url "https://www.youtube.com/watch?v=VIDEO_ID" \
    --output futsal_dataset/images/all

# Extract from specific time range
python scripts/extract_frames.py \
    --url "https://www.youtube.com/watch?v=VIDEO_ID" \
    --output futsal_dataset/images/all \
    --start 00:05:00 --end 00:15:00
```

8.3. Step 2: Annotate Frames

Each extracted frame must have bounding boxes drawn around every visible player (class 0) and the ball (class 1). Three annotation tools cover different use cases. Roboflow is the recommended entry point: it is web-based, free at the tier required for a dataset of this size, supports collaborative annotation, and exports directly in YOLO v11 format. LabelImg is a lightweight desktop alternative (pip install labelImg) suitable for offline use. CVAT provides the most professional feature set - interpolation across frames, keyboard shortcuts, and team management - but carries a steeper setup cost appropriate for larger multi-session projects.

Each label file is a plain text document sharing the filename of its corresponding image. Every line represents one object in YOLO normalised format: class_id x_center y_center width height, where all coordinates are expressed as fractions of the image dimensions in the range [0, 1]. Two classes are defined: 0 for player and 1 for ball. Annotation quality is the single most important determinant of fine-tuned model performance; tight boxes around full player bodies and consistent marking of even partially occluded balls consistently outweigh increases in dataset size in downstream metrics.

8.4. Step 3: Split into Train and Validation Sets

Once annotation is complete, the frames and their label files must be partitioned into training and validation subsets. The `split_dataset` script shuffles the annotated pool and distributes files according to the specified proportions, placing outputs into the directory structure expected by Ultralytics and auto-generating the `futsal_dataset.yaml` configuration file that YOLO training requires. An 80/20 train-validation split is used by default.

```
python scripts/split_dataset.py --dataset futsal_dataset --split 0.8
0.1 0.1
```

8.5. Step 4: Train YOLO11

Training is invoked through the same `train.py` script used in the main pipeline, passing the auto-generated YAML file as the dataset argument. The nano variant (`yolo11n`) is recommended for initial iteration because its small parameter count keeps training times under ten minutes on a GPU and under forty minutes on CPU, allowing rapid feedback on annotation quality before committing to a larger model. The medium variant (`yolo11m`) is appropriate once the dataset exceeds approximately three hundred annotated frames and a GPU is available.

```
# Nano model - recommended starting point
python scripts/train.py --data futsal_dataset/futsal_dataset.yaml

# Medium model with GPU (300+ annotated frames)
python scripts/train.py --data futsal_dataset/futsal_dataset.yaml \
    --model yolol1m.pt --batch 32 --epochs 100

# Resume an interrupted run
python scripts/train.py --data futsal_dataset/futsal_dataset.yaml -
-resume
```

Training outputs are saved under `runs/detect/futsal_train/`. The file `best.pt` in the `weights` subdirectory contains the checkpoint with the highest validation fitness score and is the model weight file used for inference. The file `results.png` plots training and validation

loss curves alongside mAP50 and mAP50-95 across epochs, providing the primary diagnostic for overfitting and convergence.

8.6. Step 5: Deploy the Fine-Tuned Model

Substituting the fine-tuned weights into the pipeline requires only a single configuration change. The `model_name` parameter in the `Config` object is updated to point to `best.pt` in place of the generic `yolo11n.pt`. No other pipeline code requires modification: the YOLO interface is identical for pre-trained and fine-tuned variants.

```
config = Config(model_name="runs/detect/futsal_train/weights/best.pt",
yolo_conf_threshold=0.3)
```

8.7. Minimum Dataset Size and Training Time Estimates

The table below gives indicative frame counts and training durations for each YOLO11 model variant. These figures assume 50 epochs on hardware representative of what is accessible at semi-professional club level. Starting with the nano model on 100–200 frames is strongly recommended: it validates the full pipeline end-to-end in under an hour before any significant annotation investment is made.

Model	Min. Frames	GPU Training Time	CPU Training Time
yolo11n (nano)	100–200	5–10 min	30+ min
yolo11s (small)	150–300	15–30 min	2–4 hours
yolo11m (medium)	300–500	1–2 hours	8+ hours

8.8. Tips for Better Results

Frame diversity is the most reliable lever available at small dataset sizes. Frames should span different areas of the court, both teams in different configurations, varied lighting conditions, and ball positions including ground-level, in-flight, and partially occluded instances. A dataset of 150 maximally diverse frames will generalise better than 300 frames sampled densely from a single two-minute passage of play.

On the training side, early stopping (`--patience 10`) prevents overfitting on small datasets by halting training when validation mAP stops improving. Increasing the epoch count beyond 50 is only warranted when the validation loss curve is still visibly declining at epoch 50; if it has plateaued, additional epochs add computational cost without accuracy gains. If GPU memory is exhausted, halving the batch size is always preferable to switching to CPU.

8.9. Common Issues

If FFmpeg is not found, verify the installation by running `ffmpeg -version` in a terminal; if the command is not found, the binary is not on the system PATH and must be added manually or reinstalled. If training reports that no labels are found for images, confirm that label filenames exactly match image filenames (frame_00001.jpg paired with frame_00001.txt) and that label files are in the `labels/all/` directory, not co-located with images. Out-of-GPU-memory errors are resolved by reducing the batch size, typically halving it until the error disappears. Persistently low detection accuracy after training nearly always indicates an annotation quality problem rather than a model capacity problem; reviewing a sample of label files in a visualisation tool is the most productive diagnostic step before adding more data or switching to a larger model.

9. Analysis of Results

9.1. Validation of the Qualitative Research

The six-interview qualitative research produced a finding of remarkable consistency: across all six professionals, at all competitive levels studied, the pattern was identical. Every professional acknowledged the value of data. None had a system that provided data reliably, automatically, and affordably. This consistency rules out idiosyncratic attitudes and points to a structural cause: the absence of appropriate tools. The research thus validates the central premise of this thesis - the data deadlock in futsal is an infrastructure problem, not a cultural one, and it is solvable.

9.2. Technical Pipeline Performance

The five-sprint pipeline produced measurable results at each stage. Sprint 1 established reliable automated video access for public YouTube content. Sprint 3 achieved detection rates above 85% per frame under favourable lighting and camera conditions. The dual-tracker architecture - PlayerTracker (Hungarian + EMA) for people, BallTracker (Kalman) for the ball - attempts to separate the distinct motion characteristics of each entity type. However, real-world performance is considerably more variable than laboratory conditions suggest, and the results must be interpreted with care.

Sprint 4's 6-point least-squares homography reduced positional error at pitch corners by 42.9% relative to the 4-point baseline (from 0.91m to 0.52m) under controlled testing conditions. Sprint 5's three-layer filtering reduced accumulated distance inflation from 8–22% to under 3% in the tested clips. These figures should be understood as best-case results: they were obtained on clips recorded under adequate lighting with clearly contrasting jersey colours. The system's performance degrades substantially as recording conditions worsen.

9.3. Alignment Between KPI Framework and Pipeline Output

Of the nine KPIs in the framework (Chapter 5), five are directly computable from the pipeline's positional output: distance covered, top speed, sprint count, possession time, and duel time. These are precisely the KPIs most consistently cited by professionals as operationally useful and currently unmeasured. The system delivers exactly what the empirical research identified as most needed.

9.4. Limitations

The system's limitations are more extensive than a summary of positive results might suggest, and intellectual honesty requires stating them plainly. Three categories of failure were observed consistently during testing and are unresolved at the time of submission.

First, team classification fails in a non-trivial proportion of clips. The K-Means classifier operating on HSV shirt colours produces incorrect team assignments when jerseys are perceptually similar (dark blue versus black, white versus light grey), when pitch-side advertising boards introduce large chromatic regions, or when video compression reduces colour fidelity. In several tested clips, one or both teams were partially or fully swapped, rendering all team-differentiated KPIs - duels, possession, duel time - unreliable for those segments. The classifier has no self-correction mechanism once the initial centroid assignment is made on the first frame.

Second, player ID stability is heavily dependent on recording brightness and contrast. In clips recorded under uneven or low indoor lighting - a common condition in regional futsal venues - YOLO's confidence scores drop below the 0.30 threshold for multiple players per frame. When this occurs, the PlayerTracker loses the track and assigns a new ID to the player when they re-appear in a subsequent frame. The practical consequence is that a single physical player may accumulate KPIs under two or more track IDs across a match segment, splitting and understating their true metrics. This is not a

rare edge case: it was observed in the majority of clips recorded in venues without professional broadcast lighting.

Third, ball tracking does not function reliably at the resolution of typical semi-professional match footage. The futsal ball is small relative to the frame, and YouTube video compression aggressively eliminates the high-frequency detail that distinguishes it from the pitch surface or from boots in close contact. The BallTracker’s Kalman filter can predict position for up to five frames in the absence of a detection, but in practice the ball is undetected for far longer stretches than this, particularly during play close to the ground or in crowded areas. Possession and duel KPIs that depend on ball position are therefore unreliable in these conditions. Ball tracking should be considered a feature of the system that works intermittently rather than consistently. The 6-point calibration is performed manually, introducing operator-dependent error. An inaccurate calibration propagates systematic positional bias across all KPIs derived from board-space coordinates. Automated calibration using detected pitch line intersections is identified as the most important near-term engineering improvement.

10. Conclusions

10.1. Verification of the Central Hypothesis

The central hypothesis is partially confirmed with important qualifications. The evidence demonstrates that an AI pipeline built on standard open-source computer vision tools can, under favourable conditions, transform fixed-camera futsal match footage into structured positional data and derive from it a meaningful set of per-player KPIs. The pipeline architecture - YOLO detection, Hungarian-algorithm tracking, 6-point homography, K-Means team classification, Kalman ball tracking - is technically sound and functions as designed. However, the qualifier “under favourable conditions” is not a minor caveat. As documented in Section 8.4, the system fails in ways that are consequential for practical use: team assignment errors corrupt team-differentiated KPIs, brightness-dependent ID loss splits player metrics across multiple tracks, and ball tracking is unreliable at the resolution of typical semi-professional footage. These are not theoretical limitations but observed behaviours across the clips tested during development. The hypothesis as stated - that the system is deployable to any semi-professional club - is therefore an overstatement. It is more accurate to say that the pipeline demonstrates feasibility under controlled conditions and constitutes a validated first iteration that requires further engineering before it can be relied upon as a production tool.

10.2. Key Contributions

The qualitative research established, with high consistency across six professionals, that the data gap in Spanish futsal is structural rather than cultural. The dual-tracker architecture (PlayerTracker + BallTracker) separates the physically distinct motion characteristics of players and ball, yielding higher KPI accuracy than a unified tracking approach. However, when using the tracker in YouTube streams, the ball is hardly ever detected. The 6-point least-squares homography reduces corner positional error by 42.9% relative to the standard 4-point approach. The three-layer spatial filter reduces distance inflation from up to 22% to under 3%.

10.3. Future Work

The following directions are identified for future development, ordered by expected impact on practical usability. Roster management and player identity resolution. The current system assigns anonymous numerical track IDs that do not persist across clips and carry no semantic meaning. A Roster module would allow a club to define a squad list - each entry consisting of player name, shirt number, and optionally a reference image - and then associate each track ID with a specific named player either automatically (via number detection or appearance matching) or through a simple post-processing UI where the analyst confirms or corrects assignments. This would transform the system's output from "Player 4 covered 3.2 km" to "Carles Puig covered 3.2 km" - the difference between a technical curiosity and a genuinely usable coaching tool. The Roster module is identified as the highest-priority feature for the next development phase. Automatic event-triggered video cutting.

A second high-priority feature is a video segmentation tool that automatically extracts short clips from the match footage when a defined event condition is met in the KPI stream. Conditions could include: a player exceeding the sprint threshold, a duel event occurring in a defined pitch zone, a possession change, or a player entering or exiting the penalty area. The output would be a library of tagged short clips that a coach can review directly, without scrubbing through full-match footage. This feature would dramatically accelerate post-match analysis workflows and aligns directly with the use cases described by the professionals interviewed in Chapter 6. Automated pitch calibration. Manual calibration introduces operator-dependent error and creates friction in the deployment workflow.

A future version of the system should detect pitch line intersections automatically using Hough line detection or a dedicated line-segmentation model, computing the homography without user input. This would reduce setup time from approximately 30 seconds to zero

and eliminate a category of systematic error. Integration of SAM 3 (Meta AI, May 2026). The recently released Segment Anything Model 3 produces pixel-accurate segmentation masks that could replace both the YOLO detection stage and the K-Means team classification stage, enabling direct jersey-colour extraction at pixel level and eliminating the bounding-box cropping approximations that cause team assignment errors under difficult lighting. Integration is contingent on GPU availability at the target deployment hardware level. Extension to event-level KPIs. The current system is limited to positional KPIs derivable from trajectory data. Extending to pass completion rate, pressing events, and shot classification would require either a dedicated event detection model or a hybrid approach combining automated position tracking with lightweight manual annotation for event types. A full web-based dashboard - accepting a YouTube URL and returning a complete post-match analytics report - would complete the user experience.

10.4. Unexpected Findings and Personal Reflection

This section steps outside the formal structure of the thesis to offer a more personal account of what the project actually felt like to build, what surprised me, and what I would think differently about if I were starting today.

When I started this project, I had a clear image in my mind of what the end result would look like: a smooth overhead tactical board showing ten coloured dots moving fluidly across a pitch, with names attached, distances accumulating in real time, and the whole thing running off a YouTube link. Something you could show a coach at half-time. That image was, in retrospect, significantly more ambitious than what a single-semester thesis project with a fixed laptop and no GPU budget could realistically deliver. I did not fully appreciate that gap at the outset - and I think that initial optimism, though ultimately humbling, was also what kept the project moving forward when the reality turned out to be more complicated.

The biggest unexpected finding was not technical - it was human. I assumed the qualitative interviews would be the easier half of the research: talk to six coaches, confirm that the problem exists, move on to building the solution. What I did not expect was how consistently the conversations revealed not just a lack of tools, but a sophisticated and frustrated analytical intelligence. Jordi Barrero knew exactly which metrics would change his selection decisions. Martí Puig had already built a manual system sophisticated enough to inform half-time substitutions at the national team level. Roger Bonet had a goal taxonomy that any data scientist would find rigorous. These were not people waiting to be introduced to the concept of data; they were people who had been trying to work with data for years with entirely inadequate means. That realisation - that the demand was already there, fully formed - sharpened the project considerably. It made the technical work feel genuinely consequential rather than a demonstration exercise.

On the technical side, the unexpected finding was the inverse: I expected the hard parts to be the mathematics - the homography derivation, the Kalman filter, the Hungarian assignment. Those turned out to be manageable, even satisfying. The genuinely difficult parts were the ones that sound trivial: getting reliable team colour classification to survive a single substitution, keeping player IDs stable across a badly-lit corridor of a regional sports hall, and convincing YOLO that the ball exists when it is two pixels wide at 720p. The gap between “working in the demo clip” and “working on footage from a cup game a Tuesday evening” was the real lesson of this project. Robustness, not novelty, is where the work actually lives.

Am I satisfied with the result? Partially, and honestly. The pipeline architecture is sound; the mathematical reasoning behind each component is rigorous; the platform for the clubs is easy to use; the qualitative research is, I believe, genuinely original in the futsal context. What I am not satisfied with is the gap between what the system does in controlled conditions and what a coach would actually need it to do in the field. I knew that gap

existed before writing the limitations section - I lived with it throughout development - but writing it down formally in Section 8.4 forced a clarity that is both uncomfortable and necessary. A system that fails on team classification in a third of real-world clips is not ready for deployment. Calling it a validated first iteration is the honest framing, and it is the one I stand behind.

What I learned that I did not expect to learn: that the combination of domain knowledge and technical build is genuinely rare, and genuinely useful. Most of the computer vision literature on sports tracking is written by researchers who have never coached a team and do not know what a pivot does in a 4-0 system. Most futsal coaches who think about analytics have no idea what a homography is. Sitting at the intersection - being able to interview Jordi Barrero about duels and then go home and implement a Kalman filter - gave this project a coherence that I think would have been harder to achieve from either side alone. That is perhaps the thing I am most proud of, and the learning I will carry furthest beyond the thesis itself.

References

- Anderson, C., & Sally, D. (2013). *The Numbers Game: Why Everything You Know About Football is Wrong*. Penguin Books.
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R. C., Mellor, S., Schwaber, K., Sutherland, J., & Thomas, D. (2001). Manifesto for agile software development. Agile Alliance. <https://agilemanifesto.org>
- Caetano, F. G., de Oliveira, M. J., Marche, A. L., Nakamura, F. Y., Cunha, S. A., & Moura, F. A. (2015). Characterization of the sprint and repeated-sprint sequences performed by professional futsal players, according to playing position, during official matches. *Journal of Applied Biomechanics*, 31(6), 423–429.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business School Press.
- Drucker, P. F. (1954). *The Practice of Management*. Harper & Row.
- García-Unanue, J., Pérez-Gómez, J., Giménez, J. V., Felipe, J. L., Bohórquez, M. R., Gallardo, L., & Sánchez-Sánchez, J. (2017). Physical and physiological demands of futsal referees. *Journal of Human Kinetics*, 60(1), 117–126.
- Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1–3), 185–203.
- James, B. (1982). *The Bill James baseball abstract*. Self-published.
- Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2), 83–97.
- Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company.
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of IJCAI*, 674–679.

- Meta AI. (2026). Segment Anything Model 3 (SAM 3). Meta AI Research. <https://ai.meta.com/sam3>
- Redwood-Brown, A., O'Donoghue, P., Nevill, A., & Seward, C. (2019). The impact of performance analysis on football. *International Journal of Sports Science & Coaching*, 14(2), 195–202.
- Schwaber, K., & Beedle, M. (2002). *Agile Software Development with Scrum*. Prentice Hall.
- Shi, J., & Tomasi, C. (1994). Good features to track. *Proceedings of CVPR*, 593–600.
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. *Proceedings of ECCV 2022*.

Annex A: Interview Instrument

The following topic areas and guiding questions structured the semi-structured interviews:

A.1 Current data practices: What data, if any, does your club currently collect? How is it processed and used?

A.2 Desired metrics: If you had access to any performance metric, which three to five would be most valuable? Why those specific metrics?

A.3 Existing tools: What tools have you tried or considered? What were their limitations?

A.4 Barriers: What are the main obstacles preventing systematic data collection at your club?

A.5 Solution reaction: [Pipeline described]. Does this address the barriers you identified? What would need to be true for you to use this?

Annex B: Sprint Retrospective Template

Field	Content
Sprint number and title	E.g., Sprint 3: YOLO + Dual-Tracker Architecture
Guiding question	The specific technical question the sprint was designed to answer
Approaches explored	All approaches evaluated, including those discarded and their formal failure modes
Decision made	The design decision taken and the mathematical/empirical justification
Limitations identified	Constraints or failure modes identified during testing on real footage
Backlog input	What the sprint revealed that should be addressed in subsequent sprints

Annex C: Full Technology Stack

Component	Library / Tool	Version	Role
Video access	yt-dlp	2024.x	YouTube stream URL extraction
Video processing	OpenCV (cv2)	4.9	Frame decode, rendering, calibration UI
Object detection	Ultralytics YOLO	v11 nano	Player and ball bounding box detection
Player tracking	Custom (greedy Hungarian)	-	Persistent ID assignment, EMA smoothing
Ball tracking	Custom (Kalman 1D×2)	-	Ball position smoothing and occlusion prediction
Perspective transform	OpenCV findHomography	4.9	6-point least-squares homography computation
Team classification	scikit-learn KMeans	1.4	HSV-based team and referee separation
Numerical computation	NumPy	1.26	Array operations and matrix maths
Data export	pandas + json	2.1	CSV (KPIs) and JSONL (positions) output
Dashboard visualisation	Streamlit	1.35	Interactive web-based post-match dashboard

Annex D: Configuration Parameter Reference

Parameter	Default Value	Description
model_name	yolo11n.pt	YOLO model file (nano variant for CPU deployment)
board_width / board_height	700 / 350 px	Tactical board dimensions (40m × 20m, 2:1 aspect)
player_class_id	0	COCO class ID for 'person'
ball_class_id	32	COCO class ID for 'sports ball'
yolo_conf_threshold	0.30	Minimum YOLO confidence for detection retention
min_players_for_kmeans	6	Minimum visible players required to train team classifier
sprint_threshold_ms	5.0 m/s	Speed threshold for sprint classification
duel_radius_m	1.5 m	Maximum inter-player distance for duel classification
possession_radius_m	1.5 m	Maximum player-ball distance for possession classification
max_frames_missing	10 frames	Frames before PlayerTracker deletes a lost track
max_distance_for_assignment	150 px	Maximum matching distance in PlayerTracker (board space)

ball_max_distance	100 px	Ball discontinuity threshold (track reset if exceeded)
kalman_process_noise q	0.01	BallTracker model uncertainty
kalman_measurement_noise r	4.0	BallTracker sensor uncertainty