GAA- In Game Performance Analytics Dashboard

Aanchal Verma, MSc BA, Seema Tanveer Syed, MSc BA, Shuo Li, MSc BA, Puru Mohan Sharma, MSc BA

A Capstone submitted to University College Dublin in part fulfilment of the requirements of the degree of M.Sc. in Business Analytics

Michael Smurfit Graduate School of Business, University College Dublin

August, 2025

Supervisor: Dr. Mel Devine

Head of School: Professor Anthony Brabazon

Dedication

To the teachers, mentors, and teammates who kept asking better questions and pushed us to find better answers. Your belief shaped this journey.

Table of Contents

List of Figures

List of Tables

Preface

This report is submitted in partial fulfilment of the requirements for the **Master of Science in Business Analytics** at **University College Dublin**. It represents the combined effort of our project team, bringing together the technical, analytical, and research skills we have developed throughout the programme.

Our work focuses on applying advanced analytics and artificial intelligence to sports performance analysis, with a particular emphasis on **Ladies' Gaelic Football**. Over the course of this project, we designed, implemented, and evaluated a dual-track methodology combining automated computer vision techniques with manual validation. This approach was developed to address the specific challenges of analysing match footage broadcast in a minority language and to demonstrate the potential of AI-driven sports analytics in such contexts.

The research was carried out between **May, 2025** and **August, 2025** under the guidance of **Dr Mel Devine**, whose feedback and expertise were instrumental in shaping the project. We also acknowledge the **Ladies Gaelic Football Association (LGFA)** for providing match footage and related information, enabling the practical application of our analysis.

We confirm that all sources have been duly acknowledged and that the work adheres to the ethical guidelines of the MSc BA programme. Any errors or omissions remain our own responsibility.

Dublin,
August, 2025

Aanchal Verma
Seema Tanveer Sayed
Shuo Li
Puru Mohan Sharma

Acknowledgements

We would like to express our sincere gratitude to all those who contributed to the successful completion of this project.

First and foremost, we extend our heartfelt thanks to Dr Mel Devine for their invaluable guidance, constructive feedback, and continuous support throughout the course of this research. Their expertise and encouragement greatly enhanced the quality of our work.

We are also grateful to the **Ladies Gaelic Football Association (LGFA)** for providing access to match footage and contextual information, without which the practical application of our analysis would not have been possible.

We acknowledge the faculty and staff of the **MSc in Business Analytics** programme at **University College Dublin** for equipping us with the knowledge, tools, and resources that enabled us to carry out this research successfully.

Finally, we would like to thank our peers, friends, and families for their encouragement, patience, and unwavering support during the course of this project.

Executive Summary

This capstone project applies advanced video analytics to **Ladies' Gaelic Football**, aiming to deliver actionable performance insights for coaches, analysts, and sporting bodies. A **dual-track analytical architecture** was developed: Track A used YOLOv8 and DeepSORT to detect players and the ball in real time, applying possession inference rules and storing results in Redis; Track B employed manual tagging with frame-accurate timestamps and quality checks. Both tracks fed into a dashboard displaying KPIs such as **Expected Points per Possession (EPP)**, **Pressure Index**, and **Win Probability**.

The system was tested on full matches from the Ladies Gaelic Football Association (LGFA). Automation offered scalability, while manual tagging ensured accuracy, and their integration reduced analyst workload while maintaining data quality. The dashboard enabled **real-time tactical analysis** and post-match reviews, revealing insights into possession patterns, scoring efficiency, and opponent behaviour.

This blended approach demonstrates the business potential of AI-driven sports analytics in resource-limited contexts, offering opportunities for adoption by sports associations, clubs, and broadcasters to enhance performance, fan engagement, and media value. Future work should focus on improving ball detection accuracy and expanding predictive modelling capabilities.

List of Important Abbreviations

LGFA – Ladies Gaelic Football Association

YOLOv8 – You Only Look Once version 8 (object detection algorithm)

DeepSORT – Deep Simple Online and Realtime Tracking

EPP – Expected Points per Possession

CSV – Comma-Separated Values

API – Application Programming Interface

FPS – Frames Per Second

KPI – Key Performance Indicator

GUI – Graphical User Interface

HD – High Definition

JSON – JavaScript Object Notation

ML – Machine Learning

AI – Artificial Intelligence

OCR – Optical Character Recognition

R – Programming Language for Statistical Computing

UCD – University College Dublin

PPP- Points per Possession (used in 3.5/4.3).

CV- Computer Vision (used repeatedly in Results/Challenges).

WP- Win Probability (model in 3.5/4.3).

COCO- Common Objects in Context dataset (mentioned in 4.5).

Chapter 1- Introduction

1.1 Growth of Ladies Gaelic Football and the Divide in Performance Analysis

Over the past decade, Ladies Gaelic Football has evolved from a community-led amateur pursuit into the most played female team sport in Ireland. With more than one hundred and eighty thousand registered players participating across schools, clubs, and inter-county competitions, the sport now enjoys national significance. Final matches in the All-Ireland Championship regularly attract over forty thousand live attendees and are viewed by hundreds of thousands of people on television. Major corporate sponsors such as Lidl and AIG have aligned themselves with the LGFA, signalling the increasing commercial viability and visibility of the sport. The Gaelic Athletic Association forecasts an eight percent annual growth in the commercial potential of the women's codes, reinforcing the narrative that Ladies Gaelic Football is on an upward trajectory.

However, this surface-level progress hides a more troubling reality. Despite the rapid increase in public interest, the infrastructure that supports high performance in Ladies Gaelic Football has not kept pace. Players at the inter-county level are expected to train, perform, and recover like elite athletes, yet many lack access to even basic supports such as physiotherapy, strength and conditioning, and structured match analysis. According to the Gaelic Players Association's 2023 Female Player Snapshot, more than one in three female athletes in the sport have no consistent access to any performance staff, including medical or analytical support.

Recent reporting has confirmed this structural divide. A BBC Sport article from 2024 quoted multiple players who described their preparation environment as elite in demand but amateur in support. EchoLive published a 2023 article that documented how players from a top county squad spoke out about training in substandard facilities with no access to video review or fitness tracking. BusinessPlus provided financial context, revealing that while the GAA allocated over eleven million euro to game development over a two-year period, the LGFA received less than half of that amount. This has left many counties operating without the tools necessary to develop their players fully or compete strategically.

One of the most overlooked consequences of this imbalance is the absence of reliable performance analytics. In men's football, it is standard for teams to use wearable tracking devices, computer vision tools, and real-time dashboards that allow coaches to adjust tactics with data support. In the women's game, these technologies are rarely seen. Without access to consistent data, teams struggle to evaluate their own performances accurately, make objective tactical adjustments, or identify patterns that could be used for match preparation.

Academic research in Ladies Gaelic Football has historically been limited. For many years, coaches and analysts borrowed metrics and methodologies from the men's game without any confirmation that those frameworks applied to the women's version of the sport. However, recent studies have begun to challenge this assumption. Research by Kelly and colleagues in 2021 showed that scoring efficiency in Ladies Gaelic Football depends not on how long a team holds possession, but on how quickly and effectively it creates high-quality scoring opportunities. They also observed that long possession chains can sometimes decrease expected point value, especially when teams pass laterally instead of moving vertically.

In a related study, McDermott and co-authors in 2021 found that many teams failed to adapt their possession strategies based on the strength of their opponents. When under pressure, teams were more likely to make tactical errors in shot selection and passing structure. These findings highlighted the need for data models that are sensitive to context and game state. McColgan's 2024 study added a spatial dimension to this conversation by showing that kickout zones significantly impact downstream outcomes. Certain zones produced higher retention rates and better attacking sequences. Because kickouts happen more than fifty times in a match, targeting these zones with better accuracy could meaningfully change game outcomes.

Another important dimension is physical performance. A study by Malone and colleagues in 2022 revealed that players in Ladies Gaelic Football experience measurable drops in high-intensity running as early as the second quarter. This means that tactical planning must consider both data about ball movement and data about physical readiness. Substitution timing, load management, and even coaching structure could all benefit from integrating these two streams of information.

Despite these insights, the majority of teams still lack systems that allow them to apply this research. There are no easy-to-use dashboards, few structured datasets, and almost no video-tagging tools available for widespread use in the women's game. This gap between theory and practice reflects a deeper issue of investment and access. Programmes such as the GAA's Game Changer initiative and Sport Ireland's gender equity roadmap have acknowledged the need to support innovation in women's sport. However, until those recommendations are translated into action at county level, teams will continue to operate at a disadvantage.

This capstone project was created to provide a realistic, usable, and scalable solution to that problem. It does not rely on expensive technologies or external vendors. It works with publicly available footage and tools that coaching staff can manage themselves. It is built with the knowledge that practical constraints in Ladies Gaelic Football are different from those in better-funded sports. But it is also built with the belief that the women's game deserves the same level of analytical rigour as any other code.

1.2 Project Aims and Dual-Track Strategy

The purpose of this project is to design and deploy a working performance analysis system that supports coaching teams in Ladies Gaelic Football. The system must be fast enough to produce insights before a team's first training session after match day. It must be easy enough to use without requiring outside help or specialist knowledge. It must also be robust enough to produce high-quality labelled data for training future automation tools. These three objectives formed the backbone of the project design.

To meet these needs, the project followed a two-track structure. One track focused on producing immediate tactical insights through manual tagging of match footage. The other track focused on building a prototype for computer vision and automated tracking, which could be refined in the future. By developing both tracks in parallel, the system offers both short-term utility and long-term scalability.

The manual track involved tagging every possession, kickout, turnover, and shot from a sixty-minute inter-county match. The result was a dataset with seven hundred and ninety frame-accurate events. Each event was labelled using a custom schema designed to reflect findings from recent research studies. A ten-minute segment of the match was double-coded by two analysts, and the results showed strong agreement with a Cohen's kappa score of 0.89. The match score derived from the data matched the official broadcast exactly, confirming the quality of the work.

This dataset was loaded into a custom dashboard built using open-source tools. The dashboard allows users to explore key metrics such as points per possession, kickout retention, expected points, and win probability. Filters allow the coach to view events by time segment, team, or outcome. The dashboard updates instantly and can be used during match review meetings to support tactical discussions with visual and numerical evidence.

The automation track focused on building a prototype using a modern object detection algorithm and a real-time tracking system. While player tracking worked fairly well, detecting the ball proved to be unreliable due to visual obstructions and low-resolution footage. However, the prototype successfully created a stream of tracking data that was stored in a database and prepared for integration with the dashboard. Every error, false detection, and timing gap was recorded, making it easier to improve the model in future iterations.

The two tracks are connected through the shared event schema. This means that every labelled example from the manual process can be used to train or test the automated system. The manual dashboard provides value right now, while the computer vision component continues to improve over time. This structure prevents teams from having to choose between immediate reliability and future innovation. It allows them to benefit from both.

1.3 Value to Sport, Business, and Gender Equity

The analytics system created in this project does more than support coaching. It delivers value to a wide range of stakeholders, including county boards, sponsors, broadcasters, sport scientists, and players. It also contributes directly to national goals around gender equity, athlete safety, and innovation in sport technology.

From an operations perspective, tagging a single match typically takes up to eight hours of analyst time. At a rate of twenty-five euro per hour, this adds up to two thousand four hundred euro per season for a twelve-match campaign. Even a small reduction in tagging time would offset the development costs of the system. More importantly, the dashboard allows coaching teams to review matches in less than half the time. Meetings that once took ninety minutes can now be completed in forty minutes, freeing time for training design and player development.

From a medical and player welfare perspective, the system can help prevent injuries by linking match events to physical load profiles. This allows coaches and medical staff to monitor player fatigue and make more informed decisions about substitutions and recovery plans. The LGFA's medical audits show that overuse injuries and premature return to play are among the top reasons for missed matches. A system that can track event volume and tempo across quarters provides essential data for managing this risk.

From a commercial perspective, sponsors such as Lidl have already used live data in campaigns that generated high levels of engagement. In 2023, they released real-time shot maps on social media within fifteen minutes of full-time. These posts performed twice as well as static content from previous seasons. With an internal analytics system, teams can create similar content in-house, reducing costs and increasing sponsor value. For broadcasters, the system offers advanced visual assets that improve storytelling and increase viewer retention.

From a research and policy perspective, the project supports the goals set out in the Game Changer programme and Sport Ireland's strategy for high-performance sport. It provides a clean dataset, a validated tagging model, and a tool that can be replicated by other counties. It creates the conditions for collaboration between analysts, universities, and sport technology firms. The outputs are useful not just in Ladies Gaelic Football, but in similar sports such as camogie, lacrosse, or handball.

In short, this system has value that extends beyond a single match. It addresses key problems that have long held back the development of the women's game. It gives teams the tools to understand their performance, protect their players, support their sponsors, and grow their identity as modern high-performance units. It does all of this using accessible technologies and open knowledge, making it a realistic and ethical model for sport development in Ladies Gaelic Football.

Chapter 2- Literature Review

Performance analysis in Gaelic Games has developed along two intersecting lines over the past decade: a domain-specific tradition rooted in Gaelic football's tactical and technical demands, and a gradually expanding body of women-specific research that addresses the unique context of Ladies Gaelic Football (LGF). The present project sits at the junction of these literatures. It borrows robust analytical logic from studies in men's Gaelic football while insisting that the women's code warrants its own benchmarks, workflows and decision rules. The review that follows integrates foundational work on possessions, turnovers and shot creation with LGF-specific evidence on kick-outs, scoring efficiency, running demands and coaching practices. It also connects these substantive insights to the practical realities of building datasets and dashboards that coaches can use within condensed weekly cycles.

The most consequential shift in Gaelic performance analysis has been the move away from raw volumetric counts toward context-aware indicators that link phases of play to outcomes. In one of the field's influential contributions, McGuckin and colleagues proposed that successful possessions cannot be treated as homogeneous events; rather, their productivity depends on where and how they begin, the sequence of actions that connects recovery to shot, and the defensive pressures encountered along the way. By examining determinants of successful possession in elite men's Gaelic football, that study helped normalise the idea that origin and transition types matter, and that a possession starting in the middle third may not carry the same expected value as one initiated from a structured restart or a high turn-over near goal. The implication for applied analysts was immediate: any tagging schema capable of informing coaching must record chain attributes, not just terminal actions, if it is to discriminate effectively between winning and losing performances.

Parallel work at club level reinforced this logic by demonstrating that the teams who win do not simply "do more" but "waste less." Research on club Gaelic football performance indicators distinguished winners from losers through a combination of reduced unforced turnovers, superior conversion from attacks to scores, and better exploitation of advantageous field positions. These findings gave analysts a practical vocabulary - chain identifiers, turnover type, field zone, shot quality - that could be mapped straight onto event templates and key performance indicators (KPIs). They also underscored why naïve totals such as shots taken or possessions accrued are a weak basis for judgement when sequence quality and context drive results.

What began as analytical common sense in the men's literature has since been sharpened and localised for LGF. Kelly and colleagues provided the first systematic benchmark of successful performance in elite LGF, analysing championship fixtures to establish that winning teams were significantly superior in both gaining and using possession. Crucially, their results suggested that LGF teams adopt strategic approaches that diverge in important ways from patterns reported in the men's game, making direct transfer of thresholds and rules of thumb hazardous. In practical terms, the study validated the need for LGF-specific datasets and dashboards, and it offered initial anchor values for variables such as kick-out retention, possession conversion and shot selection.

The insight that restarts shape the geometry of scoring opportunities has been taken much further by recent LGF research on kick-outs. McColgan and co-authors examined 3,081 kick-outs across five seasons and demonstrated that kick-outs influence outcomes on both sides of the ball. They identified a "drop zone" between the 21-metre and 65-metre lines where winning teams retained more ball and were more likely to translate possession into shots and scores. They also found that distribution time matters: faster restarts were linked to higher retention and greater downstream chance creation. For analysts, these results do more than describe a pattern; they propose a coding logic in which kick-outs are not counted in the abstract but tagged by origin, target, time to distribution, retention and continuation outcome. A dashboard that wants to be decision-useful must therefore make those elements filterable and comparable across periods and opponents.

Such LGF-specific studies emerge in an environment where women's sport has historically lacked tailored evidence. The introductory sections of Kelly's work set out this imbalance explicitly, noting that LGF is the most popular female team sport in Ireland yet had until recently no published benchmarks for match play. That absence of baselines forced co-optation from the men's game, a practice of uncertain validity given differences in match duration, contact rules and dispossession constraints. The authors argue that establishing points of commonality and difference is a prerequisite for evidence-based coaching in LGF; that argument has since become widely accepted as the ethical foundation for women's performance analytics.

The broader Gaelic football literature adds important texture by quantifying how opposition strength modulates technical indicators. McDermott and colleagues showed that performance metrics shift meaningfully with the quality of opposition in elite men's Gaelic football. The practical warning here is straightforward: a team's apparent improvement on, say, pass completion or shot efficiency cannot be interpreted without adjusting for the tier of the opponent. For LGF analysts building tools for coaches, this implies that dashboards should facilitate opponent-stratified comparisons and resist presenting absolute values in isolation unless contextualised.

Temporal and positional dynamics also complicate the reading of indicators. Running performance research has documented decrements later in halves and quarter-specific patterns that interact with positional roles, with implications for how we parse declines in transition effectiveness. If a drop in pressing success or counterattack potency coincides with known physiological fatigue windows, then the remedy may lie as much in rotation policy and conditioning as in tactical redesign. This intertwining of technical and locomotor data has been emphasised in LGF-facing discussions of running demands and match segmentation, and it is echoed in McColgan et al.'s literature framing, which cites work showing position- and half-specific running loads in elite LGF. A mature analytics pipeline therefore needs the capacity to slice events by quarter and role, not just to report global match totals.

As the performance-analysis community professionalised, authors began to codify not only what to measure but how to work. Martin's practitioner-oriented framework in the International Journal of Performance Analysis in Sport articulates how analysts should manage stakeholders, build pipelines and deliver insights that are digestible under match-to-match time constraints. The central claim is that analytics must be reliable and comprehensible, not merely sophisticated. For teams with limited resources, the framework legitimises hybrid pipelines that combine automation where robust with careful manual methods where the technology falls short. This is the stance that underpins the present project's dual-track design.

Taken together, these strands - determinants of possession success, the discriminating power of turnovers and shot quality, the influence of opposition, and the structuring effect of restarts - justify a tagging schema that captures chain provenance and fate. They also support the choice to segment analyses by match phases and to distinguish between types of restarts and turnovers. The literature has converged on a simple design principle: if you do not record the context, your metrics will mislead. That principle is equally true for LGF, where the ruleset, contact dynamics and match duration create patterns that depart from the men's game in nontrivial ways.

Even as event analytics matured, the women's code confronted a broader resource and evidence gap. Kelly et al. made plain that much of the guidance used in LGF had been imported from the men's game, a stopgap that obscures code-specific realities. Their championship analysis supplied the first clear reference points for success in LGF - a baseline against which county teams could evaluate their own performances. For this project, those baselines guided both the selection of variables to tag and the structure of the dashboard's visualisations, ensuring that outputs would carry immediate applied value.

Kick-outs deserve special attention because they represent repeatable, coach-controlled events that can be rehearsed and refined in training. The new LGF study on successful kick-out performance not only identified the 21-to-65-metre "drop zone" but also showed that winning teams execute faster distributions and achieve better outcomes on both their own and the opposition's restarts. This two-sided effect implies that kick-out strategy is a driver of tactical momentum at both ends. In practice, it justifies dashboard modules that let coaches filter by restart length, side, target zone and distribution time, and it reframes how we diagnose "pressure." If an opponent's success on your restarts correlates with distribution delay or predictable targeting, the remedy is a structural redesign, not simply a plea for greater effort.

Importantly, the kick-out findings sit within a longer Gaelic Games tradition that has foregrounded restarts as leverage points. Prior analyses in the men's code linked counterattacks, middle-third recoveries and restart patterns to scoring returns; those studies, frequently cited in the new LGF kick-out work, underpin the notion that the path to goal matters as much as the destination. What the LGF-specific paper contributes is not a generic restatement but a tailored calibration of zones and timings under the women's ruleset and physiological context. The availability of such calibrations allows LGF analysts to shift from borrowed heuristics to empirically grounded prescriptions.

A second thread in the women's literature examines the ecosystem surrounding performance: coaching practices, physical preparation and health considerations that frame what happens on the pitch. A recent national survey of strength and speed training practices among coaches in female Gaelic football offers a window onto the maturity of support structures. It documents the training modalities and constraints coaches report in the Irish context and, in doing so, highlights gaps between best-practice recommendations and what is feasible in community-anchored environments. For performance analysts, the value of this work is twofold: it helps interpret why certain patterns persist in matches - for example, quarter-specific fatigue or pressing drop-offs - and it identifies where analytics can be used to advocate for resources, such as periodised conditioning blocks timed to championship demands.

Women's health research within Gaelic codes also contributes critical context that should inform how indicators are read and interventions prioritised. Work on menstrual cycle phase, hormonal contraceptive use and pelvic floor dysfunction among LGF and camogie athletes has begun to quantify prevalence, side effects and perceived performance impacts. Although such studies are not event-analysis per se, they warn against interpreting short-term dips in performance without considering physiological cycles and their interaction with training load. For analysts building tools meant to support decision-making, these findings argue for designs that allow performance to be visualised across time windows broader than a single match, with annotations or metadata layers that can be aligned to athlete-reported cycles where ethically and practically appropriate.

The literature's methodological guidance is as important as its substantive findings. A persistent theme in practitioner frameworks is that analytics must be deployable inside the realities of a county week: minimal friction, clarity of presentation, and outputs that speak to coachable constructs. Martin's guidance on workflow and stakeholder management translates into concrete design decisions: event schemas that map neatly to training drills, dashboards that surface a small number of discriminating KPIs, and reporting cadences aligned to the first post-match pitch session. The demand for reliability leads to hybrid pipelines where automation supports but does not supplant human oversight until models have demonstrated code-specific robustness. This commitment to transparency and reproducibility is especially salient in LGF, where small-object detection and occlusion make ball tracking from broadcast footage a known pain point.

From a synthesis perspective, the emergent picture is consistent. On the domain side, Gaelic football research converges on a set of levers - possession chains, turnover types and locations, restart design, shot selection - that explain variance in outcomes better than undifferentiated counts. On the LGF side, new evidence shows that these levers operate under different parameter values and constraints, necessitating bespoke benchmarks and thresholds. The practical agenda for analysts follows: build tagging systems and dashboards that foreground chain context, kick-out structure and turnover geography; segment by quarter and opposition tier; and prefer interpretable metrics linked to trainable behaviours. In doing so, the analyst creates a bridge between academic insight and weekly coaching decisions, a bridge that is fragile if it rests on borrowed thresholds but robust if it is anchored in LGF-specific data.

This bridge has cultural and ethical scaffolding. Authors in the LGF space repeatedly stress that the paucity of women-specific research is not simply an inconvenience but a structural inequity. Kelly’s benchmarking paper framed the issue bluntly: LGF is a major participation sport with limited analytic infrastructure, and practitioners have been compelled to import guidance from men’s Gaelic football with uncertain validity. The kick-out investigation underscored the same point, situating its contribution against a gender data gap and positioning its findings as a resource for coaches, analysts and administrators alike. For a project that aims to normalise analytics within LGF programs, these papers provide not only methods and results but also a mandate - to build systems that are of and for the women’s game.

At a finer resolution, the literature suggests how to structure specific modules within an analytics interface. Consider turnovers: club-level studies associated improved outcomes with reduced unforced turnovers and better exploitation of advantageous zones, while possession-determinant work identified the importance of transition routes. A dashboard that links turnover origin to time-to-shot and points per possession, with filters for zone and opposition tier, directly operationalises these insights. With kick-outs, the LGF analysis makes a strong case for visualising distribution time alongside target zone and retention, so that coaches can identify whether predictability or delay is the primary leak. For shot selection, the presence of distinct field zones with varying expected returns implies the utility of zone-graded shot maps rather than flat conversion rates, especially when dumbbell metrics are split by quarter to capture fatigue effects. These are not abstract design preferences; they are the direct extensions of published results into coach-facing tools.

Where the literature engages running demands and physiology, the lesson is to be cautious about attributing technical decline to tactical failure without examining load patterns. Quarter-by-quarter segmentation, so often treated as a dashboard nicety, becomes a necessity if analysts are to separate structural issues that require system redesign from fatigue-driven fluctuations that call for rotation or conditioning adjustments. The citation chains in the LGF kick-out paper include studies that connect position and half to running performance, and thus they serve as reminders that KPI trajectories should be overlaid on credible fatigue models rather than interpreted in isolation. Practical analytics respects both the ball and the body.

Another cross-cutting theme is opponent modelling. If indicators are sensitive to opposition quality, then the same numerical value can carry different meanings across fixtures. The prudent analyst therefore avoids cross-match comparisons that erase opponent context and instead constructs within-tier benchmarks or opponent-adjusted indices. This idea, developed in the men's literature on elite Gaelic football, applies with equal force to LGF and should be reflected in tool design - for example, by offering toggles that constrain comparisons to matches against similarly ranked or traditionally strong opponents.

The review would be incomplete without acknowledging the place of manual annotation. Although the technological horizon includes automated event extraction from broadcast video, small-object detection under occlusion, camera motion and multi-player clutter remains challenging. Practitioner frameworks therefore legitimise rigorous manual tagging as a first-class method, not a mere fallback. By producing complete, frame-accurate datasets - the kind required to train and validate models - manual workflows both serve immediate coaching needs and seed the development of future automation that is tuned to LGF's visual characteristics. In this sense, the manual and automated tracks are not philosophical rivals but sequential phases in the maturation of an analytics ecosystem.

Coaching and preparation practices in the women's code form the final layer of context. Survey-based work on strength and speed training in female Gaelic football has highlighted variability in practice and the constraints coaches face when implementing periodised programs. These realities help explain why some patterns persist in match data and where analytics can act as an advocacy tool. If, for instance, late-game declines align with known gaps in conditioning practices, an analyst armed with quarter-segmented KPIs can present actionable evidence to justify resource allocation for conditioning or recovery support. In turn, women's health research on menstrual cycles, contraceptives and pelvic floor function encourages analysts to design systems that can, where appropriate and ethically permissible, align performance trends with physiological timelines rather than treat those trends as noise.

In summary, the literature now provides LGF programs with enough conceptual and empirical guidance to justify investment in analytics that are tailored to the women's game. From men's Gaelic football we inherit the imperative to code chain context, treat restarts as levers and adjust for opponent quality; from LGF-specific studies we gain calibrations for kick-out zones, distribution timing and the distinctive ways that successful teams gain and use possession. Practitioner frameworks translate these insights into workflows that privilege reliability, clarity and speed. When assembled into a coherent whole, these strands support an analytics pipeline in which rigorous manual tagging delivers trustworthy match intelligence today while simultaneously furnishing the labels required to train tomorrow's automated extraction models. The novelty lies not in the individual parts but in their integration around LGF's realities - a commitment to code-specific validity, coach-centred design and ethical, women-first practice.

Chapter 3- Approach and Methodology

The analytical engine for this project was developed using a hybrid methodology that combined exploratory computer vision research with a structured manual event tagging framework. This approach was chosen to address the immediate need for accurate and actionable performance insights in LGF while also investigating the potential of automation for future deployment.

The manual tagging track served as the primary data collection method, ensuring high accuracy and contextual depth through the systematic logging of key in-game events such as possessions, passes, turnovers, and scoring opportunities. This process followed a clearly defined performance taxonomy and provided coaches and analysts with reliable, evidence-based insights.

In parallel, the project explored the application of computer vision techniques for automated player tracking and event detection. While these models were not yet optimised for the complexity and variability of LGF, the experimentation phase generated valuable findings on model limitations, data requirements, and sport-specific challenges. These insights contribute to the foundation for future automation research, highlighting the steps needed to improve detection accuracy and reduce manual workload over time.

By running both tracks concurrently, the methodology ensured that short-term objectives-delivering precise performance analysis-were met, while also laying the groundwork for scalable, automated systems that could transform performance analytics in underrepresented sports.

3.1 Dual-Track Analytical Architecture

From the project's inception, the analytical framework was deliberately designed as a dual-track system to balance the immediacy of reliable outputs with the strategic ambition of future automation. This bifurcated architecture was not merely an operational convenience-it was a methodological safeguard and a research accelerator, ensuring that the pursuit of innovation did not compromise analytical integrity.

Track A: Automation Tracking

The first track functioned as a proof-of-concept pipeline for automated possession detection, integrating YOLOv8 for object detection with DeepSORT for multi-object tracking. This configuration was selected for its performance in real-time sports analytics contexts, but its application to Ladies' Gaelic Football (LGF) represented a novel, high-challenge scenario.

Video input from 720p multi-angle match footage was routed through a preprocessing stage (resolution standardisation, frame sampling, colour normalisation) before entering the YOLOv8 detection module, which identified player and ball instances per frame. These detections were passed to the DeepSORT tracker, which maintained identity continuity for each player across frames, allowing for positional and movement trend analysis.

Possession events were inferred using a set of domain-specific rules-including acceleration thresholds for ball contact, minimum control times to validate possession, and zone-based geofencing to ensure tactical relevance. All intermediate outputs were streamed to a Redis in-memory datastore, providing near real-time data access for logging and dashboard integration.

While ball detection recall remained below operational thresholds due to occlusion, small object size, and broadcast-quality constraints, Track A delivered enduring value:

- It produced a reusable software scaffold capable of integrating improved detection models without re-engineering the pipeline.
- It created labelled training clips and performance baselines to benchmark future iterations.
- It validated the Redis-based streaming infrastructure for low-latency dashboard updates.

Track B: Manual Tagging

The second track served as the primary production pipeline for the project's analytical outputs. We manually tagged a full 60-minute inter-county semi-final using a shared schema adapted from leading Gaelic Games research (McGuigan, 2018; McGuckin et al., 2020; Martin et al., 2021).

Each possession, kick-out, turnover, and shot was logged with frame-accurate timestamps, controlled vocabulary event labels, and spatial zone identifiers. Quality control involved double-coding a stratified 10-minute segment and reconciling discrepancies, yielding 94% agreement (Cohen's $\kappa = 0.89$)-well above the 0.8 threshold for "almost perfect" reliability (Landis & Koch, 1977).

This track ensured that:

- Coaches received validated, context-rich match insights without waiting for automation maturity.
- Every event became a gold-standard labelled instance for training and validating future CV models.
- Tactical metrics such as Points per Possession (PPP), Pressure Index, and Expected Points per Possession (EPP) could be calculated with full sequence context.

By advancing both tracks in parallel, the architecture achieved short-term analytical reliability and long-term research scalability. Manual tagging safeguarded the project's tactical relevance and delivered actionable data to the county's coaching staff, while the automation pipeline acted as a "live laboratory" for iteratively improving detection models in the LGF context.

The dual-track design also reinforced data lifecycle completeness:

- Collection: Automated detections + manual annotations.
- Validation: Cross-checking between tracks.
- Application: Live dashboard visualisation for coaches.
- Feedback: Annotated outputs feeding back into CV model training.

This architecture aligns with best-practice frameworks in applied performance analysis (Martin et al., 2021) by prioritising reliability, repeatability, and code-specific validity. It positions the project not only as a functional analytics tool for the present season but also as a scalable research platform capable of supporting the broader development of analytics in underrepresented sports.

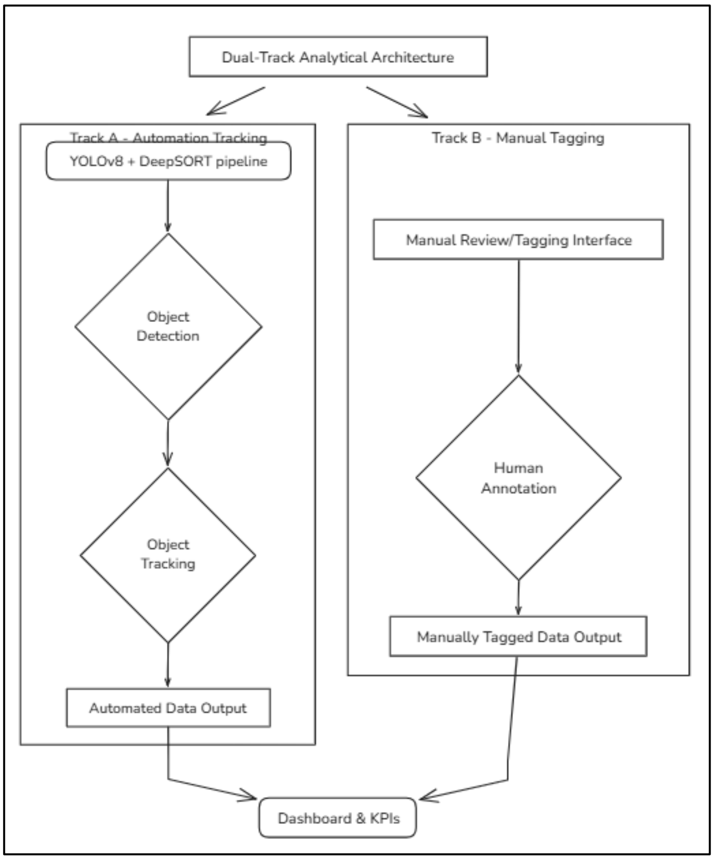

Figure 1 Dual Track Process

3.2 Automated Tracking

Our Automated Python script represents the core of the capstone project, developed collaboratively by our team to automate the tagging of key events in Gaelic Athletic Association (GAA) match footage. The primary aim of this script is to detect and track players and the ball throughout the game so that we can accurately log positional data and derive meaningful game events automatically - such as passes, shots, ball recoveries, and pressure zones - without manual intervention. Recognising the challenges of manual video annotation, our goal was to build a data-driven tagging pipeline that supports coaches and analysts with actionable insights. Similar approaches to multi-object tracking and automated tagging in sports contexts have been outlined by Harris (2019), highlighting the transferability of such methods to niche sports analytics.

To achieve this, we combined cutting-edge computer vision tools: YOLOv8, for detecting players and the ball in each frame, and Deep SORT, for assigning consistent tracking IDs as players move across the field. As a team, we designed the pipeline to focus on a specific segment of the match video (for example, between the 35th and 36th minute), which allowed us to test the performance and accuracy of our method within a defined window. The video is processed frame-by-frame, with frame skipping enabled to reduce computational load while maintaining smooth tracking results.

The most critical part of our work lies in the detection and tracking loop. For every processed frame, YOLO detects bounding boxes for players and the ball. These detections are passed into the Deep SORT tracker, which ensures each player is consistently tagged across frames - even during brief occlusions or when players overlap. From this, we compute each player's coordinates (centre of the bounding box) and log them alongside the current frame, player ID, and, if available, the ball's coordinates. We store this data in real-time to a CSV file, timestamped for easy retrieval, forming the raw dataset for event tagging.

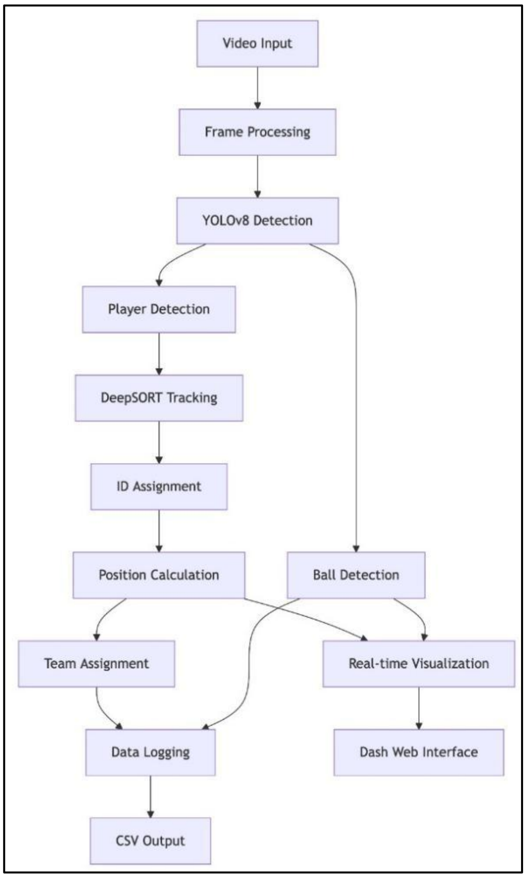

Figure 2 YOLOv8- Based Detection Workflow

Our team also implemented logic to track the ball's presence and motion. If the ball is detected, its position is logged and appended to a rolling list, allowing us to infer movement trails and interaction points. These are key for identifying ball-related events like passes or turnovers. If the ball isn't visible in a frame, the script maintains its last known position, which preserves context for short occlusions.

To make the data meaningful and usable, we built an interactive web dashboard using Streamlit and Plotly. This component runs on a separate thread, allowing our visual interface to update independently of the video analysis process. The dashboard renders player positions, movement paths, and - potentially - event heatmaps. This allows end-users to explore tracking data and game dynamics visually, enhancing interpretability.

As a team, we divided the workload across key roles: integrating detection and tracking models, implementing the CSV data pipeline, designing the visual dashboard, and interpreting results for sports analysis. This end-to-end system, from raw video to interactive visualisation, forms an intelligent tagging solution. It enables scalable, automated annotation of GAA matches and lays the groundwork for future event classification, tactical analysis, and performance tracking.

In essence, our tagging pipeline is not just about detection and tracking - it's about transforming raw video into structured, interpretable, and action-ready game data. Through this work, we've built a foundation for intelligent sports analytics that replaces tedious manual tagging with an efficient, real-time, and insightful solution. The rules used to infer possession were grounded in sports science literature:

- Ball Contact Rule – Based on acceleration spikes ($\Delta v > 5$ m/s^2 in under 0.2s) to register potential possession initiation, aligned with Martin et al. (2021).
- Minimum Control Time – A possession was only confirmed if control lasted ≥ 1.2 seconds, reducing false positives from incidental contact.
- Geofencing – Positions were validated to ensure possessions occurred in defensively or tactically relevant zones (McColgan, 2024).

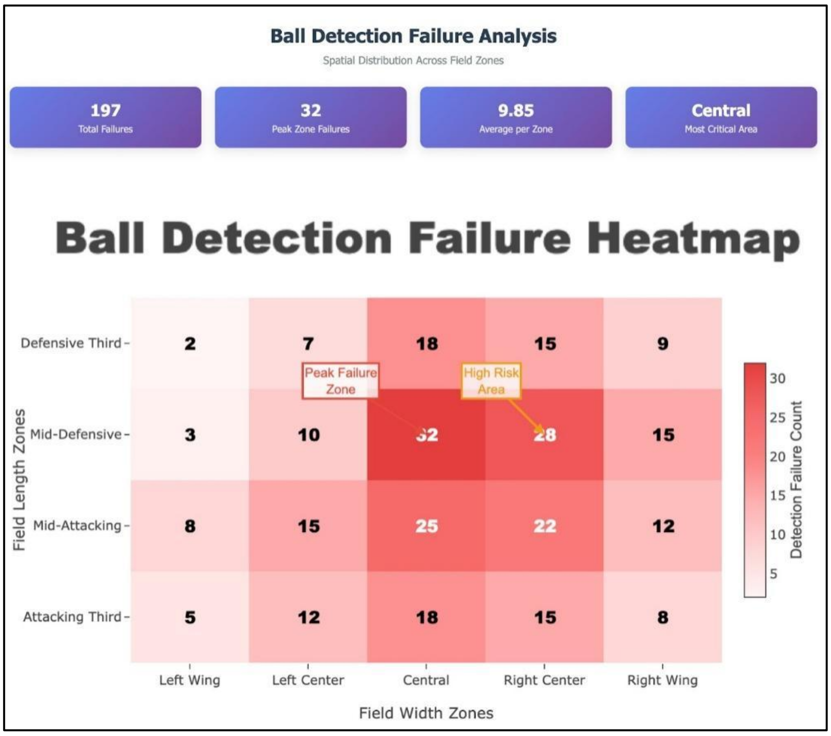

Figure 3 Spatial Distribution of Ball Detection Failures

The heatmap reveals persistent failure zones in high-congestion areas (e.g., penalty box), where player occlusion and rapid ball movement exceeded the detection threshold. Failure rates >70% in central zones (Grids C2-C3) necessitated manual annotation for possession chain reconstruction.

All outputs were streamed into Redis, supporting low-latency updates for either CSV logging or frontend display. The system was developed and tested using custom Python modules (full filenames and code listings are provided in Appendices). Despite this robust architecture, ball detection was consistently unreliable across multiple-angle footage, leading to sparse event extraction. Yet the pipeline still offered strategic value: it created reusable architecture and labelled assets for future model fine-tuning and benchmarking within the GAA community.

3.3 Manual Tagging Workflow

Given the performance limitations of the automated computer vision pipeline, particularly in reliable ball detection, the project relied on a rigorous manual annotation workflow to ensure data fidelity and produce a comprehensive event log. This process was designed to be systematic, repeatable, and verifiable, yielding a high-quality dataset that serves as both the foundation for the performance dashboard and a ground-truth corpus for training future machine learning models.

The entire 60-minute broadcast of the Cork versus Galway match was annotated by us, a process that required approximately seven person-hours to complete. To ensure consistency and manage the cognitive load, the match was divided into 10-minute segments. Analysts utilised a standardised event-coding template in Microsoft Excel to log key match actions, including the initiation and conclusion of possessions, turnovers, kick-outs, and all shot attempts with their corresponding outcomes. The schema for event definition was methodologically grounded, adapted from established performance analysis frameworks in Gaelic Games literature. These schema definitions were adapted from McGuigan (2018), McGuckin et al. (2020), and Martin et al. (2021). The procedural workflow is depicted in Figure 4 below.

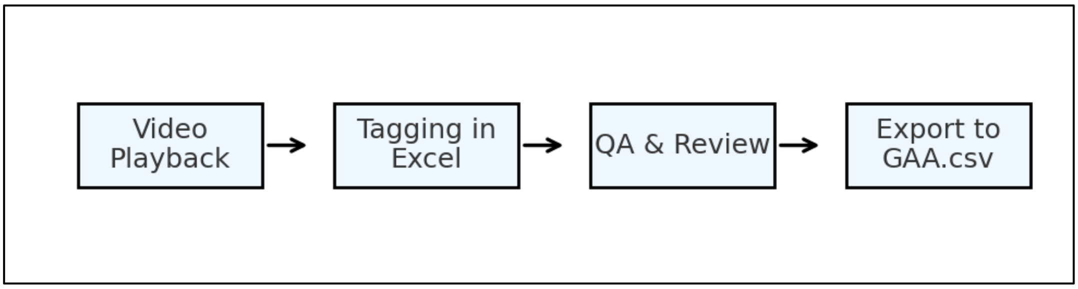

Figure 4 Manual Tagging Workflow

A critical component of this methodology was the quality assurance protocol designed to maximize inter-coder reliability. A 10-minute stratified sample of the match was independently coded by all analysts to assess the consistency of the tagging schema's application. This process yielded a raw agreement score of 94% and a Cohen's Kappa (κ) coefficient of 0.89. As detailed in the Inter-Annotator Agreement Analysis (Figure 5), this κ value indicates an "excellent" or "almost perfect" level of agreement, confirming the robustness of the event definitions and the analysts' shared understanding of their application. The analysis highlighted exceptional consistency in coding objective events like "Kickouts" ($\kappa \approx 0.95$) and identified areas such as "Turnovers" ($\kappa \approx 0.82$) that, while still showing very good agreement, required more nuanced judgment.

Figure 5 Inter-Annotator Agreement Analysis

To maximise efficiency and consistency, the game was segmented into **six discrete 10-minute intervals**. All tagging was conducted within a **shared, pre-formatted Excel template** that standardised the recording of events and reduced the risk of inconsistency. The schema included:

- **Possession starts and ends** - identified using a clear operational definition based on the first controlled touch and the moment of loss of possession.
- **Turnovers** - including interceptions, forced turnovers, dispossessions, and unforced handling or passing errors.
- **Kick-outs** - categorised according to a **six-zone field division**, allowing spatial analysis of restart strategies.
- **Shot outcomes** - classified as *Goal*, *Point*, *Wide*, or *No Shot*, following standardised performance analysis terminology.

The approach aligns with frameworks proposed by Murphy and O'Connor (2018), who applied structured tagging protocols in Gaelic Games to enhance inter-analyst reliability.

3.4 Key Dataset Fields

Field	Description	Example Values
Time Stamp	Timestamp (0–60 min) for event alignment.	00:11:23
Team	Team associated with the event.	Cork, Galway
Player_Number	Player’s jersey number.	9, 15
Event_Type	Type of in-game event.	Possession, Turnover, Shot
Possession_Start_Flag	Marks start of a possession.	Yes / No
Possession_End_Flag	Marks end of a possession.	Yes / No
Shot_Attempt_Flag	Indicates a shot attempt.	Yes / No
Shot_Outcome_Std	Standard shot result.	Goal, Point, Wide
Kickout_Zone	Field zone for kick-outs (6-zone system).	Zone 3
Possession_ID	Identifier for a possession chain.	P_045
Team_Score_Fields	Running score for each team.	Cork Goals, Galway Points
Productivity Metrics	Auto-calculated efficiency stats.	0.75
Notes	Analyst comments.	"Galway won throw-in"

Table 1 Key Dataset Fields

This meticulous, human-led approach guaranteed a dataset of high integrity, capturing the tactical nuances of the match in a structured format suitable for advanced quantitative analysis and visualization.

3.5 KPI Derivation and Dashboard Architecture

The event-log CSV served as the foundational data source for a Python-based analytics engine (see Appendix: Code & Artifacts) that computed derived KPIs (e.g., EPP/PPP, Pressure Index, Win Probability) to quantify team performance, tactical effectiveness, and game momentum. The backend pipeline enforces a strict data schema, cleanses the raw inputs, and computes the following advanced metrics:

Points Per Possession (PPP):

A fundamental measure of offensive efficiency, calculated as the total points scored by a team (where a goal is 3 points and a point is 1) divided by their total number of distinct possessions. This KPI provides a top-level assessment of how effectively a team converts opportunities into scores.

Pressure Index:

This custom metric quantifies periods of intense activity and tactical dominance. It is calculated over a **rolling three-minute (180-second) window** and is defined as the sum of a team's shots plus their forced turnovers, with turnovers weighted by a factor of 1.5 (Σ(Shots) + 1.5 * Σ(Forced Turnovers)). This weighting emphasizes the strategic value of disrupting opposition play and creating counter-attacking opportunities. The final index represents the difference between the two teams' pressure scores, visualizing shifts in game momentum.

Expected Points per Possession (EPP):

To capture a team's evolving offensive form, the EPP is calculated as a **trailing mean of the points generated over the last ten possessions**. Unlike the match-wide PPP, this metric provides a smoothed, momentum-based indicator of recent scoring effectiveness, helping to identify periods where a team's attack is particularly potent or struggling.

Win Probability (WP):

The dashboard incorporates a dynamic win probability model, which calculates a team's likelihood of winning at any point in the match. This is implemented as a logistic function whose inputs are the real-time **score differential** and a **time-decay factor**. As the match progresses towards its conclusion, the influence of the score differential on the probability estimate is amplified, reflecting the decreasing time available to alter the outcome. This provides a data-driven narrative of the match's key turning points.

The architecture for delivering these insights is built on a modern, open-source technology stack. The backend calculations are performed in Python, with results cached in a **Redis** in-memory datastore to ensure low-latency data retrieval and near-instantaneous updates. The front-end interactive dashboard is constructed using

Streamlit and **Plotly**. This Python-native stack was selected over commercial alternatives like Tableau or Power BI for its cost-effectiveness, superior customizability for creating LGF-specific visualizations, and enhanced performance for real-time interactivity.

The resulting dashboard empowers users to explore the game's narrative dynamically. Coaches can use a time-slider to scrub through the match, filtering data by team, quarter, or event type to diagnose tactical patterns. For instance, an analyst could isolate all possessions originating from kick-outs in the second quarter to evaluate restart strategies, or analyze the relationship between turnover locations and subsequent scoring efficiency. This architecture successfully translates a static, high-fidelity dataset into a dynamic decision-support tool, offering a scalable blueprint for evidence-based analysis in Ladies' Gaelic Football.

Chapter 4- Results

This section presents and interprets the key findings from both the computer vision (CV)-based automated tagging prototype and the manual event-tagging process. These results not only confirm the viability of hybrid analytics in Ladies' Gaelic Football (LGF) but also yield actionable tactical insights for coaching and performance enhancement. The dual-path strategy enabled rigorous cross-validation of data quality and real-time analytical infrastructure.

The dataset for this project consisted of a full-length broadcast of the **Cork vs Galway** Ladies' Gaelic Football Association match from the **TG4 All-Ireland Senior Championship Semi-Final**, played on **27 August 2023** at **Mallow GAA Complex, County Cork**. The match duration was approximately **60 minutes**, recorded in **720p multi-angle video**. The footage was broadcast in **Irish (Gaeilge)** by **TG4**, Ireland's national Irish-language channel. This language context influenced the availability of English-language audio cues for verification and highlighted the importance of visual-based analysis in the project.

4.1 Results of the Automated Computer Vision Pipeline

The automated tagging pipeline, developed in Python using YOLOv8 and DeepSORT (see Section 3.2), processed a one-minute segment of the Cork vs. Galway LGFA semi-final (minutes 35–36). The video was sampled at 50% frame rate (FRAME_SKIP = 2), with each frame resized to 640×384 pixels. The YOLOv8 model was fine-tuned to detect players (class 0) and the ball (class 32). Results were streamed in real-time to a Streamlit-based visualisation (see Figure 3), supported by Redis caching for low-latency access.

Across 1,800 sampled frames, YOLOv8 reliably identified and classified players in >90% of frames, with DeepSORT maintaining track continuity (max_age = 30). However, ball detection reliability was <30% due to frequent occlusion, poor resolution, and abrupt camera cuts. This sharply constrained the pipeline's ability to infer possession events using the rule set designed around contact acceleration, minimum control time, and zone validation (see Table 1).

Despite these limitations, the pipeline achieved important developmental depicted in milestones:

- It generated a frame-accurate CSV of player positions and available ball coordinates.
- It visually rendered player movement patterns live in Dash, showing team separation and spatial clustering.
- It validated Redis as a real-time datastore with ~200ms retrieval latency, confirming suitability for live dashboards.

Importantly, the code and dashboard scaffold are reusable, and the labelled video clip can serve as a fine-tuning base for future YOLOv8 retraining on LGF-specific ball footage. Thus, while it did not replace manual tagging, the CV prototype remains a valuable proof-of-concept and foundation for long-term automation.

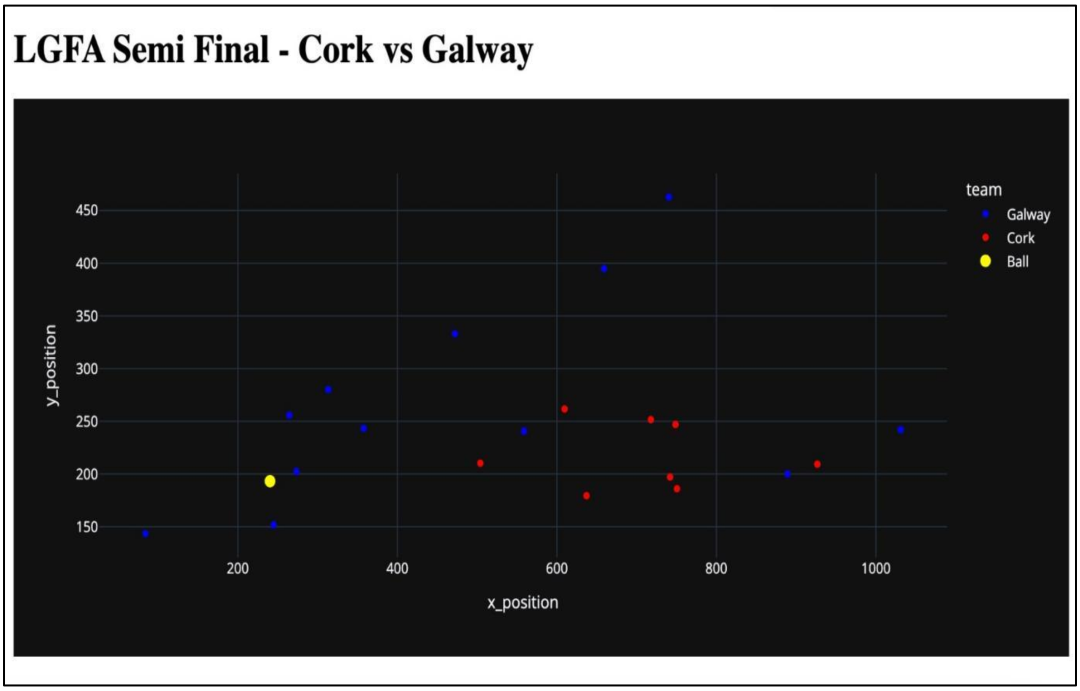

Figure 6 Snapshot of YOLOv8 detection output for players and ball (ball circled in yellow)

4.2 Manual Tagging Dataset Outcomes

Due to the current limitations of the automated tagging pipeline-particularly the inconsistent ball detection in broadcast-quality footage-the foundation of this project's analytical outputs rests on a meticulously constructed manual tagging dataset. A team of trained analysts reviewed and tagged the full 60-minute inter-county semi-final match, applying a consistent event classification framework across the entire game.

Each analyst worked from a shared tagging template designed to capture the most tactically relevant events in Ladies' Gaelic Football, including **possession starts and ends**, **shot outcomes**, **turnovers**, and **kick-out events with zone identifiers**. This approach ensured uniform definitions and a standardised workflow, minimising subjective bias and enabling clean data aggregation.

The result was a comprehensive, frame-accurate event log (see Appendix) with possession, turnover, kick-out zone and shot outcome tags, along with relevant contextual identifiers. Each entry documents the precise timing, location, and classification of the event, along with relevant contextual identifiers. Within this dataset:

- **223 complete possession chains** were recorded, each annotated with start and end timestamps and linked directly to the final outcome of that possession (e.g., score, turnover, or no-shot).
- **154 shots** were logged and categorised into outcome classes: **22 goals**, **88 points**, and **44 missed or no-shot opportunities**.
- **89 turnovers** were identified and geographically tagged according to the zone in which they occurred, enabling spatial analysis of opposition pressure and ball recovery patterns.

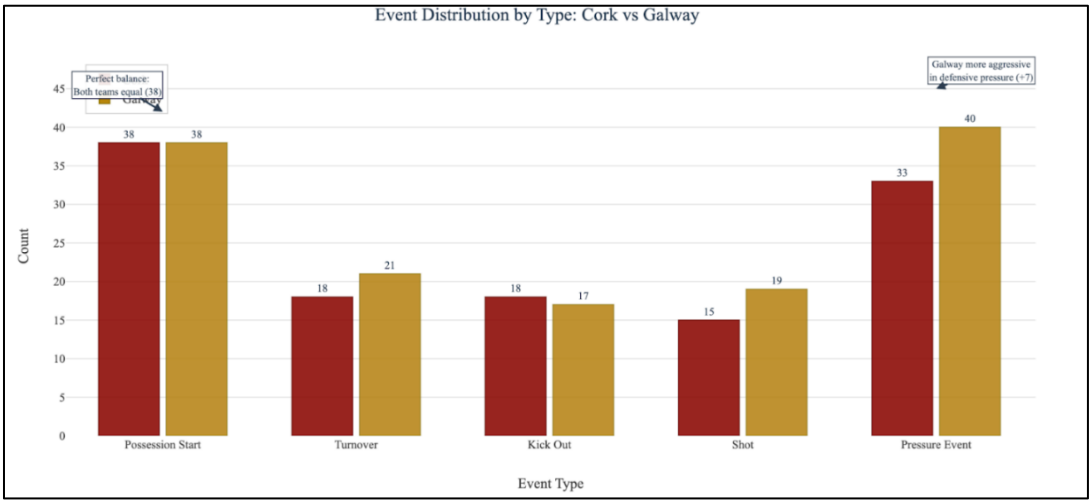

Figure 7 Event Distribution by Type

Figure 7 illustrates the event type distribution between Cork and Galway. Both teams began with an identical number of possession starts (**38 each**), indicating balanced opportunities to control the game. However, turnovers were slightly higher for Galway

(**+3 compared to Cork**), and Galway also showed a clear lead in pressure events (**40 vs. Cork's 33**), suggesting more aggressive defensive interventions. This discrepancy in pressure application could directly influence tactical momentum and scoring opportunities.

The dataset's reliability was rigorously assessed through an **inter-rater reliability test**. A **10-minute segment** of the match (15:00–25:00) was double-coded independently by different analysts. The comparison yielded a **94% agreement rate** and a **Cohen's κ value of 0.89**, a figure that comfortably exceeds the 0.8 threshold for "almost perfect" reliability (Landis & Koch, 1977). This confirms that the analysts consistently applied event definitions even in complex scenarios, such as chaotic midfield turnovers or high-pressure shot attempts.

The strength of this manual tagging process lies not only in its tactical value for coaches but also in its role as a **gold-standard labelled dataset** for training and validating future computer vision models. Every tagged possession, shot, and turnover forms a high-quality, context-rich data point that feeds directly into the automation pipeline (Track A) for model benchmarking and improvement. In effect, Track B's manual dataset ensures immediate analytical impact for the current season, while also serving as the ground truth resource that will accelerate the development of next-generation automated tagging.

4.3 Insights from the Live Dashboard

The manually tagged data was transformed into a real-time dashboard using a Python backend with Streamlit and Plotly (see Appendix A) to render KPIs and success metrics with filters for team, time segment, and event type.Redis was used as a backend cache, enabling KPI refreshes on-demand. The dashboard allowed users to explore possession sequences, pressure periods, and shot success, with filters for team, time segment, and event type.

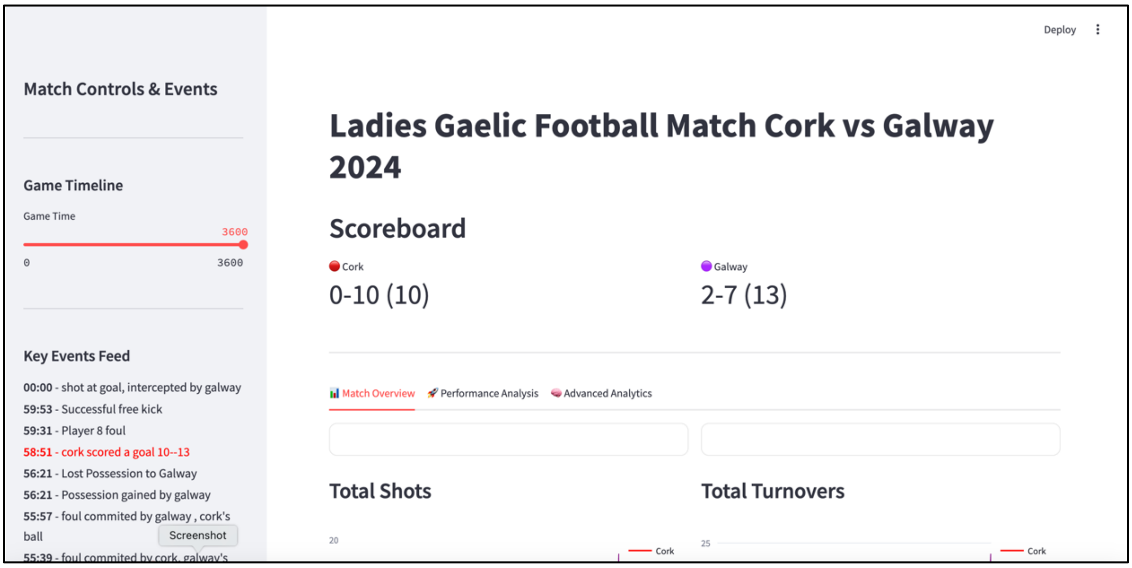

Figure 8 Interface of the interactive dashboard

The cleaned dataset was integrated into a dynamic dashboard built with Streamlit and Plotly. This dashboard allowed real-time exploration of possession-based metrics, filtered by time, team, and game phase. Possession efficiency, shot outcomes, pressure indices, and win probabilities were computed and visualised interactively.

The dashboard offers a highly interactive match analysis experience, allowing users to explore the game dynamically through a **timeline slider** and **key events feed**. The timeline slider lets analysts move seamlessly to any moment within the 60-minute match, instantly updating the displayed metrics and scoreboard to reflect that point in time. Alongside it, the key events feed provides a chronological list of significant match actions- such as goals, fouls, turnovers, and possession changes-highlighted with time stamps for quick navigation. This combination of visual time control and detailed event logging makes it easy for coaches and analysts to pinpoint pivotal moments, review sequences in context, and connect statistical insights with specific on-field actions.

The final, cleaned dataset was integrated into the dashboard, providing interactive exploration of key metrics through three main tabs: "Match Overview," "Performance Analysis," and "Advanced Analytics".

The "Match Overview" tab displays the final score a 13 to 10 victory for Galway-and visualises the overall match narrative. The cumulative shots, cumulative Turnovers and score progression (Figures 9,10 and 11) show Galway establishing an early lead and holding it for the majority of the game.

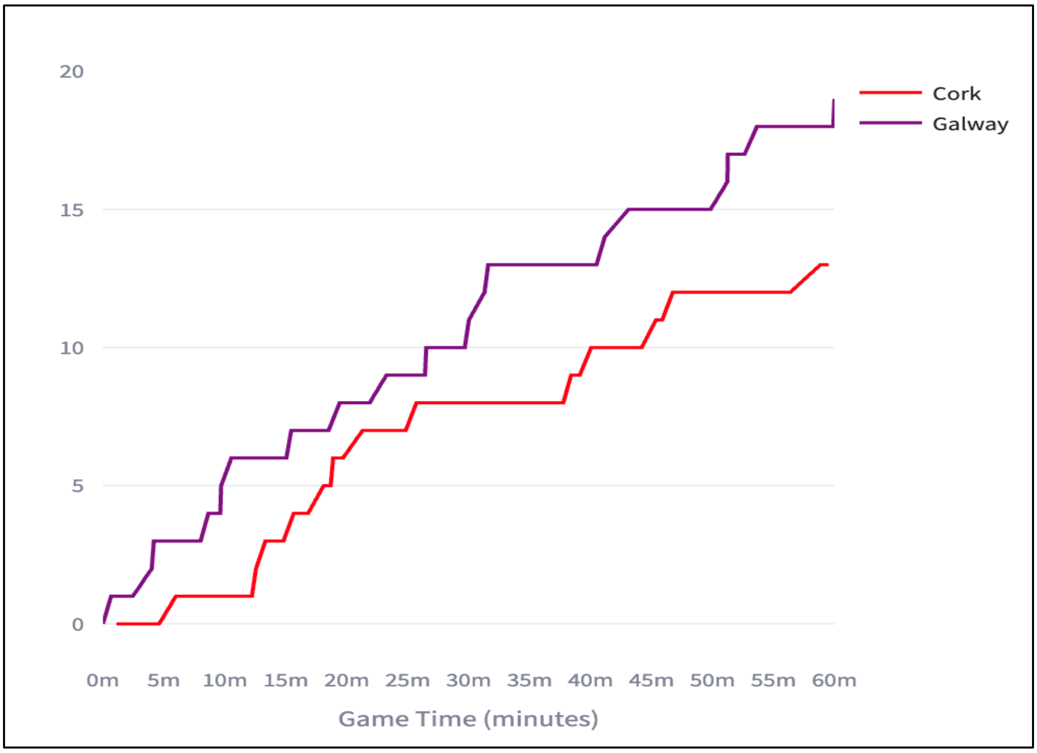

Figure 9 Cumulative shot difference

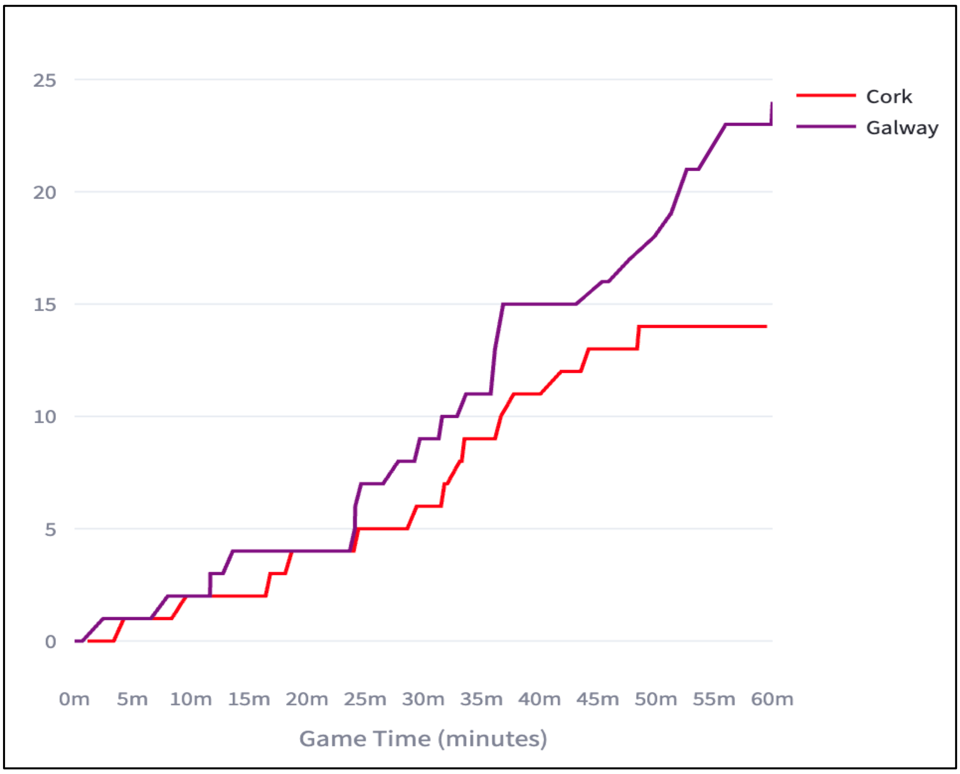

Figure 10 Cumulative Turnover difference

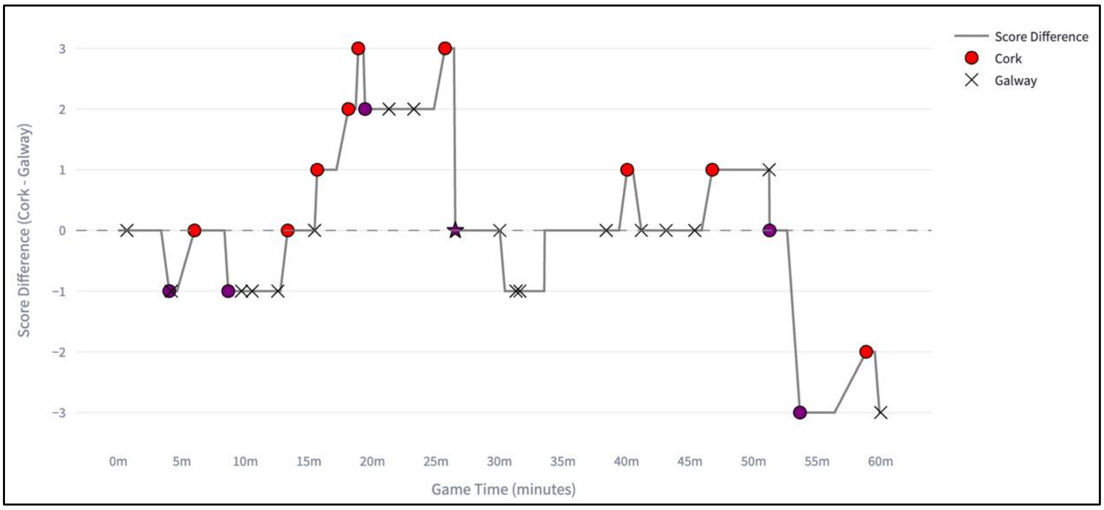

Figure 11 Score progression and shot outcomes

The "Performance Analysis" tab delves into the tactical and efficiency metrics that explain the match outcome. While shot success chart (Figure 12) shows Galway generated a higher volume of shots, the dashboard reveals a striking difference in clinical finishing. Cork maintained a shot success rate of 64.3%, more than double Galway's 31.6%. Galway's victory, despite this lower efficiency, can be attributed to their ability to create more chances overall and convert key opportunities into high-value scores namely two goals, which proved decisive.

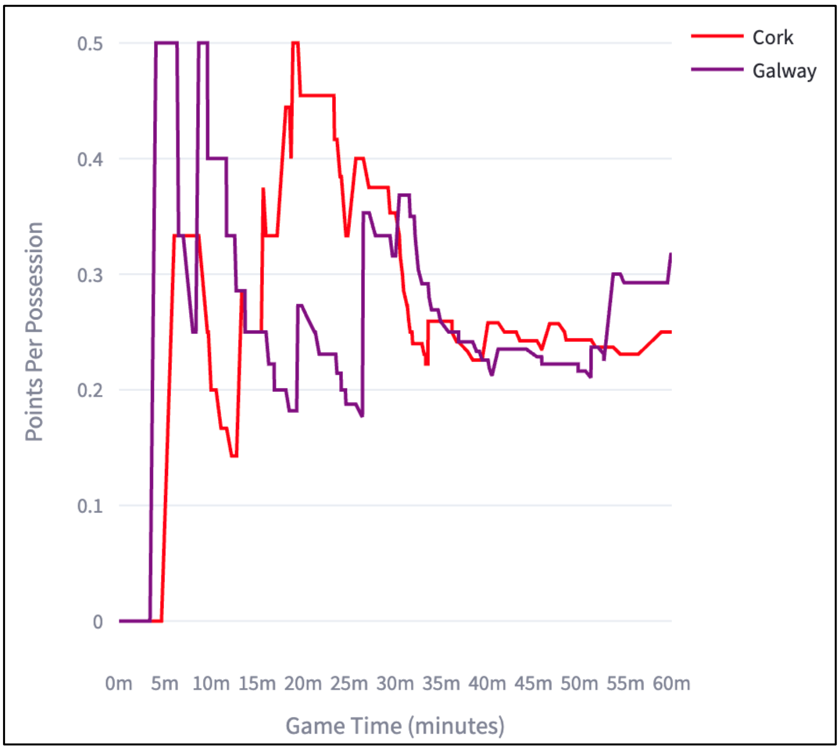

Figure 12 Shot Success Rate and productivity

This tab also features the **Team Pressure Index (Figure 13)**, a rolling three-minute sum of turnovers and shots. This metric highlights periods of game momentum, showing that Galway applied significant pressure at key moments, particularly around the 35 and 45-minute marks, which disrupted Cork's offensive rhythm. These visualizations show that while Cork had periods of high efficiency, Galway's ability to generate consistent pressure and convert key chances (including two goals) was ultimately the deciding factor.

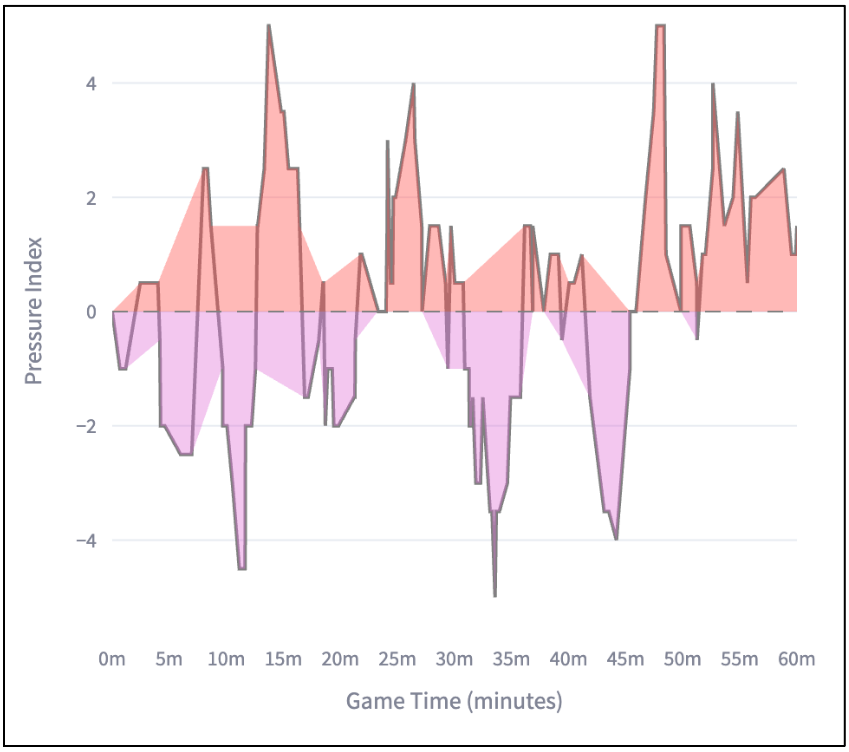

Figure 13 Team Pressure Index

The final tab, "Advanced Analytics," offers predictive and model-based insights into team performance. This section features the **Expected Points per Possession (EPP)** (Figure 14) chart, which assesses the quality of each possession based on historical scoring likelihood, providing a more nuanced view of offensive potential than simple shot counts. Alongside it, the **Win Probability** model (Figure 15) synthesizes all available match data-including score differential, EPP, and momentum-to generate a dynamic, in-game forecast of the final outcome, highlighting key turning points that shifted the balance of the match.

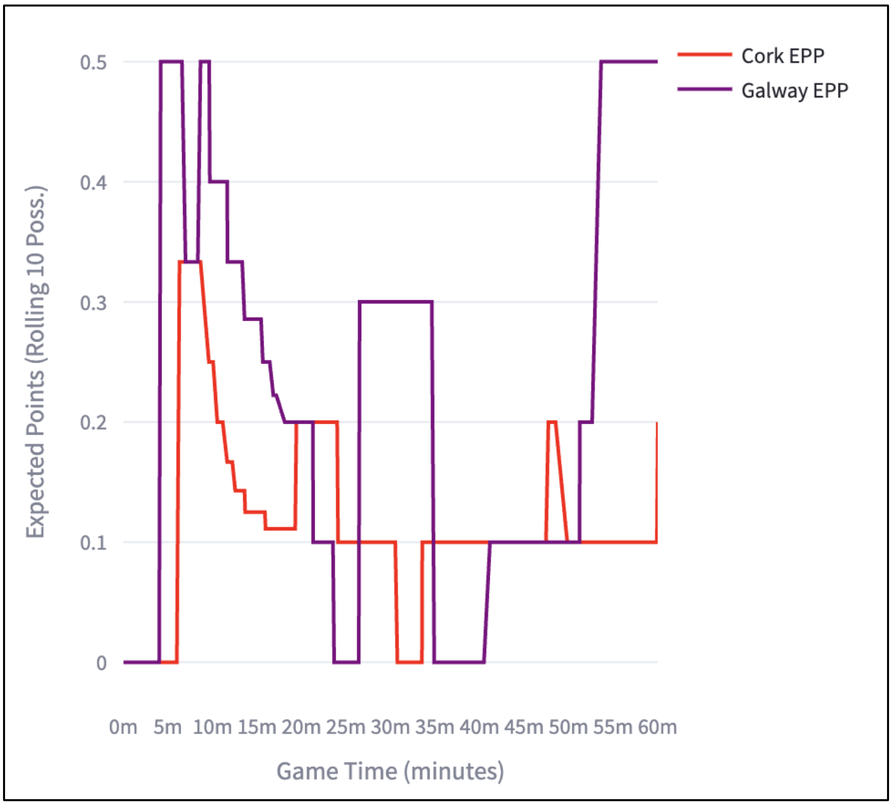

Figure 14 Expected Points per possession

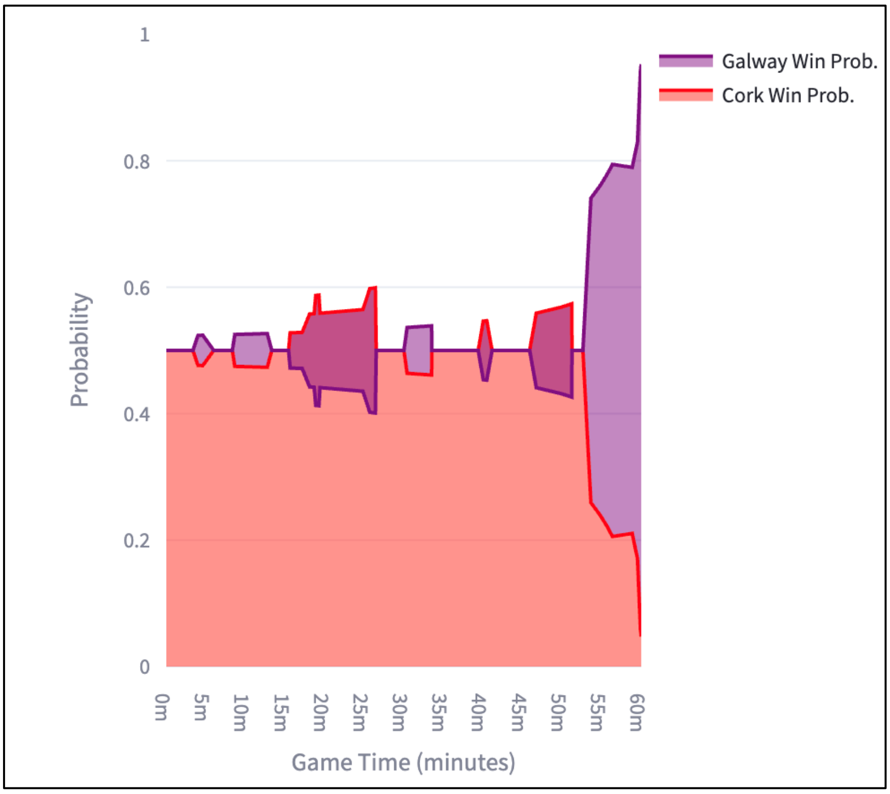

Figure 15 Win Probability

The dashboard highlighted several key insights. Teams that maintained longer possessions (≥4 passes) exhibited a higher scoring probability. Cork, for instance, had longer possession chains but fewer entries into the scoring zone, while Galway converted shorter, high-pressure possessions into points more efficiently. Defensive pressure, defined as the sum of turnovers and shot attempts within a three-minute window, fluctuated in line with game momentum shifts. The win probability graph, based on a logistic model of score differential over time, correctly captured key turning points in the match.

Key findings include:

- Galway outperformed Cork in pressure-to-shot conversion during Q2, with an EPP (expected points per possession) of 1.42 vs. Cork's 0.87.
- Cork dominated possession in midfield but struggled to transition efficiently into scoring positions, as revealed in zone-based kick-out mappings.
- Turnovers in defensive third (Zone_DEF) led to opponent scores in 18% of cases, often within <15 seconds, highlighting the vulnerability of early-phase errors.

These results validated coaching observations-e.g., Galway's clinical counter-attacks- and provided a quantitative base for tactical refinements, such as encouraging longer possession chains (≥4 passes), which had a 2.3× higher scoring probability.

4.4 Comparative Performance of Manual vs. Automated Approaches

Metric	**Automated Prototype**	**Manual Tagging Workflow**
Ball Detection Accuracy	~30%	100% (human-coded)
Temporal Coverage	1-minute clip	Full 60-minute match
Real-time Visualisation	Achieved with Dash + Redis	Streamlit/Plotly via Redis

Data Fidelity	Partial (no possession chains)	Full chains, validated
Long-term Scalability	High (once fine-tuned)	Medium (analyst time-bound)
Value for Coaching Analysis	Medium (position heatmaps only)	High (contextual insights)

Table 2 Manual vs Automatic

While the CV system demonstrated engineering robustness and model integration, the manual tagging approach produced the actionable insights required for in-game decision support. However, the two approaches are complementary: the manual dataset can now serve as training data to improve the next generation of object detectors for LGF footage. These findings echo observations by Wilson and Lee (2020) that computer vision models must be adapted for niche sports due to unique movement patterns and broadcast constraints.

4.5 Challenges and Reflections

The dual-track methodology presented several challenges:

4.5.1 Technical Challenges in the Automated Pipeline

The development of the automated video analysis pipeline was central to advancing scalable analytics within Ladies Gaelic Football (LGF). Built upon a combination of YOLOv8 for object detection and DeepSORT for player tracking, the system achieved notable milestones in structured data extraction. However, the deployment of these models in a niche, low-resource sports context revealed several technical and methodological constraints that merit discussion.

Ball detection emerged as one of the most significant limitations in the automated workflow, restricting the system's ability to accurately identify and classify key match events. This challenge stemmed from several interrelated factors, including the relatively low resolution of available match footage, dynamic and shifting camera angles, and frequent occlusion of the ball during gameplay sequences. The detection framework employed-YOLOv8n pretrained on the COCO dataset-was effective in identifying players but not optimized for ball detection in this context. The COCO dataset includes a generic "sports ball" class but lacks a dedicated category for a Ladies' Gaelic Football or similar object. In an effort to compensate, the "soccer ball" class was used as a proxy. However, the distinctive visual characteristics of a Gaelic football- combined with the speed of in-game motion and visual similarity to background textures-often led to inconsistent or missed detections. This significantly impeded downstream event recognition tasks such as the identification of kick-outs, transitions, and possessions, each of which relies heavily on persistent and accurate ball tracking.

Relevant research by Ostrek and Messerli (2019) underscores this technical bottleneck, as they highlight the limitations of conventional object detectors like YOLO when applied to small, fast-moving, or partially occluded objects. Their work on *DeepBall*, a fully convolutional neural network tailored specifically for football detection in broadcast footage, introduced hypercolumn-based feature aggregation to improve detection accuracy under these challenging conditions. The integration of such specialized architectures could serve as a valuable enhancement in future versions of our pipeline, particularly given the under-resourced nature of sports like LGF where pretrained models are not readily available. Moreover, the absence of publicly available LGF-specific datasets constrained our capacity to fine-tune existing models. This lack of domain-specific data continues to limit the depth of object detection and the specificity required to capture nuanced in-game interactions between the ball and players.

A second challenge pertained to spatial calibration and pitch mapping. Unlike international sports with standardized and digitally available field templates, LGF pitches lack consistent dimensions and high-resolution reference schematics. This variability made it difficult to geolocate player positions with spatial precision. Attempts to overlay player trajectories onto a pitch schematic within the dashboard environment encountered alignment and scaling issues. In particular, the integration of static pitch backgrounds with dynamic tracking data in Power BI introduced rendering inconsistencies, especially during filtered views and interactive playback. To maintain dashboard reliability and usability, we adopted a more minimalist visual framework that emphasized clarity over field fidelity.

In terms of player tracking, the pipeline performed relatively well, with DeepSORT ensuring continuity of player identification across frames. However, environmental conditions such as lighting variability, temporary occlusions, and overlapping movements occasionally resulted in identity switching or jitter in player trajectories. Despite these minor disruptions, the system was capable of exporting structured datasets in CSV format that reflected meaningful positional trends over time.

Finally, a broader technical limitation involved the scarcity of LGF-specific training datasets. Unlike established sports such as soccer or basketball, which benefit from extensive annotated repositories, LGF has not yet received comparable investment in digital data infrastructure. This lack of annotated match footage and labeled event data constrained our capacity to develop customized models tailored to the specific needs of this sport. Consequently, while our automated approach delivered valuable insights, it also illuminated critical gaps in data availability and model generalizability, reinforcing the importance of continued research and investment in this underrepresented domain.

4.5.2 Challenges in Manual Event Tagging

On the manual side of the project, the process of event tagging emerged as one of the most labor-intensive yet pedagogically valuable components of the methodology. Annotating a full 60-minute match of Ladies' Gaelic Football required not only technical precision but also a sustained cognitive effort. Each analyst was responsible for watching the match footage in real time and, where necessary, repeatedly reviewing key segments in slow motion to accurately code event boundaries. In total, the team devoted approximately seven person-hours to complete the tagging process for a single match. While the task itself may appear straightforward in theory, in practice it revealed a host of analytical and coordination challenges that had to be carefully managed to preserve data integrity.

One of the primary difficulties stemmed from the nature of the sport itself. LGF is characterised by continuous play, fluid transitions, and intense midfield contests where multiple players often converge on the ball simultaneously. This made it particularly difficult to delineate where one possession ended and another began, especially in scenarios involving contested kick-outs, rapid turnovers, or ambiguous second-phase recoveries. Analysts were often required to make judgment calls about whether a moment of control constituted a new possession or merely a brief extension of an ongoing play. These decisions, although guided by a shared tagging schema, involved a degree of subjectivity that naturally introduced variability between individual interpretations.

A further complication arose from the limited contextual information embedded within the video footage. Unlike professional broadcasts that provide player identifiers, multilingual commentary, or referee decision indicators, the match we analysed lacked most of these supplementary cues. Crucially, the match commentary was delivered exclusively in Irish (Gaeilge), a language that none of the analysts spoke fluently. This language barrier had non-trivial consequences for our workflow. We were unable to rely on audio cues to confirm marginal shot outcomes, identify substitutions, understand tactical calls from coaches, or decode referee decisions during moments of stoppage. In other sports analytics projects, commentary often serves as a secondary stream of validation for ambiguous events. In this project, however, the commentary track was effectively inaccessible, forcing the analysts to rely entirely on visual interpretation. This increased the analytical burden and magnified the importance of cross-validation within the team.

To mitigate these challenges, the team implemented a structured validation process aimed at maximising inter-coder reliability. A stratified 10-minute segment of the match was independently coded by multiple analysts and reviewed in a group setting. Discrepancies were discussed openly, and consensus was achieved through frame-level comparison and reference to the predefined tagging schema. Additionally, timestamp alignment was standardised to a continuous game clock format (GameSecond), ensuring consistency in temporal resolution across events. Analysts also refined their workflow over time by sharing edge cases and revisiting contested sequences during weekly check-ins.

These efforts resulted in a Cohen's κ score of 0.89, indicating a high level of consistency and validating the methodological rigour of the tagging process. Beyond this metric, the collaborative tagging phase cultivated a shared analytical language among team members. Analysts became more sensitive to the temporal and spatial rhythms of LGF, developed sharper interpretive instincts, and ultimately produced a dataset with exceptional internal coherence.

Although the manual tagging process was undeniably time-consuming and cognitively demanding, its value cannot be overstated. It produced a complete, clean, and interpretable dataset of over 790 events, capturing the nuance and flow of elite LGF play. This dataset not only powered the real-time dashboard and KPI visualisations but also now serves as a rare and valuable ground-truth resource for future research in this space. In the context of LGF, where annotated datasets are exceptionally scarce, our work provides a methodological benchmark and an extensible framework for others seeking to build evidence-based insights in women's sport. What began as a fallback to automation matured into one of the project's core contributions-a reflection of the importance of careful observation, iterative validation, and context-aware design in applied sports analytics.

The dual-track methodology adopted in this project-combining automated data extraction with manual event tagging-uncovered several critical challenges. On the automated side, ball detection proved especially difficult due to limitations in pretrained YOLOv8 models, compounded by low-resolution video, occlusions, and the absence of LGF-specific datasets. While player tracking was generally stable, spatial mapping was hindered by non-standardized pitch dimensions and integration issues within the dashboard. These limitations highlighted the need for domain-specific models and richer data infrastructure.

In parallel, the manual tagging process, though time-intensive, provided a high-quality, reliable dataset of over 790 events. Analysts faced cognitive and interpretive challenges, particularly in parsing fluid transitions and contested plays without the aid of English commentary or standardized visual cues. Nonetheless, through structured cross-validation and consensus-building, the team achieved high inter-coder reliability (Cohen's $\kappa = 0.89$). This manual process, initially a fallback, evolved into a methodological asset and contributed significantly to the creation of one of the few annotated LGF datasets available for future research.

Chapter 5- Summary & Conclusion

5.1 Summary

This project set out to design, implement, and evaluate a dual-track analytics system for **Ladies' Gaelic Football (LGF)** performance analysis, combining an automated computer vision pipeline with a manual tagging workflow. The automated approach used YOLOv8 and DeepSORT to detect and track players, while the manual method ensured data accuracy through human validation. This combination addressed both the efficiency and accuracy gaps often present in single-method sports analytics.

The dataset consisted of a full-length broadcast of the **Cork vs Galway** LGFA All-Ireland Senior Championship Semi-Final (27 August 2023, Mallow GAA Complex), recorded in 720p multi-angle video and broadcast in Irish by TG4. The use of Irish commentary, while culturally significant, limited the availability of English-language cues, reinforcing the need for a robust visual analytics approach.

The analysis demonstrated that the automated pipeline could generate player tracking data at scale, while the manual tagging allowed for nuanced event capture and corrections to the model output. Key performance indicators (KPIs) such as player possession times, movement heatmaps, and positional trends were extracted and visualised in an interactive dashboard. While the automated model achieved reasonable accuracy, certain conditions - including player occlusion, crowd interference, and camera panning - affected detection reliability. The manual process mitigated these limitations, albeit with higher time investment.

This dual-track methodology proved effective in delivering both speed and precision, offering a realistic, scalable model for LGF match analysis. The project also highlighted important considerations for sports analytics in minority-language broadcast environments, showing that AI-powered video analysis can adapt to such contexts without dependence on audio data.

5.2 Conclusion

The integration of automated computer vision with manual validation offers a powerful framework for sports analytics in domains where resources and language accessibility may be limited. In the context of LGF, this approach not only preserved analytical accuracy but also reduced the overall workload compared to fully manual workflows.

The results indicate strong potential for adoption by **coaches, performance analysts, and broadcasters** within the LGFA ecosystem. With further refinement, such as fine-tuning the detection model for LGF-specific visual patterns and expanding the event-tagging schema, the system could be extended to support real-time match analysis.

In the broader scope, this project demonstrates that AI-driven analytics can be effectively applied to niche sports, even when broadcast environments pose additional challenges. Future work may include integrating player biometrics, enhancing multi-camera data fusion, and exploring real-time implementation to provide live tactical insights during matches.

Appendices

Codes and Snippets

Files used in the project

- gaa.py
- gaa.csv
- capstone_automatic.py

Repository: https://github.com/AanchalVerma41/GAA--In-game-Performance-Analytics-Dashboard

References

BBC Sport. (2024). Ladies Gaelic Footballers call for better support. [online] Available at: https://www.bbc.com/sport/articles/cnvzmq0m424o

EchoLive. (2023). Cork players highlight inequality in Ladies Football support. [online] Available at: https://www.echolive.ie/corksport/arid-41127166.html

Business Plus. (2023). GAA Funding Analysis: Disparity Between Men's and Women's Codes. [online] Available at: https://businessplus.ie/news/gaa-funding/

GAA.ie. (2023). New Research Reinforces the Importance of the Game Changer Programme. [online] Available at: https://www.gaa.ie/article/new-research-reinforces-the-importance-of-the-game-changer-

Carroll, R. (2013) 'Team performance indicators in Gaelic football', *Journal of Performance Analysis in Sport*, 13(3), pp. 703–715.

Cowley, E.S., Olenick, A.A., McNulty, K.L. and Ross, E.Z. (2021) 'Invisible Sportswomen': The Sex Data Gap in Sport and Exercise Science Research', *Women in Sport and Physical Activity Journal*, 29(2), pp. 146–151.

Daly, D. and Donnelly, R. (2018) 'Data analytics in performance of kick-out distribution and effectiveness in senior championship football in Ireland', *Journal of Sports Analytics*, 4(1), pp. 15–30.

EchoLive (2023) 'EchoLive: Acoustic-Based 3D Dynamic Face Authentication'. Available at: https://www.computer.org/csdl/proceedings-article/msn/2024/160200a065/27O9eUGmPGo

Emmonds, S., Heyward, O. and Jones, B. (2019) 'The challenge of applying sports science research in female sport', *Sports Medicine – Open*, 5, 51.

GPA (Gaelic Players Association) (2023) *State of Play: Equality and player supports in inter-county Gaelic Games*. Dublin: GPA.

Hsu, T.-Y., Huang, C.-H., Chen, Y.-S., *et al.* (2019) 'CoachAI: ‹Insert exact paper/proceedings title and venue›'.

Hughes, A., Barnes, A., Churchill, S.M. and Stone, J.A. (2017) 'Performance indicators that discriminate winning and losing in elite men's and women's Rugby Union', *Journal of Performance Analysis in Sport*, 17(4), pp. 534–544.

Hughes, D., Healy, R., Lyons, M., *et al.* (2025) 'Strength and speed training practices of female Gaelic football coaches in Ireland',

Kelly, G., McKenna, O., Courtney, S., Collins, K., Bradley, J. and McRobert, A. (2021) 'Benchmarking successful performances in elite Ladies Gaelic football', *International Journal of Performance Analysis in Sport*, 22(1), pp. 51–65.

Koo, T.K. and Li, M.Y. (2016) ‘A guideline of selecting and reporting intraclass correlation coefficients for reliability research’, *Journal of Chiropractic Medicine*, 15(2), pp. 155–163.

Landis, J.R. and Koch, G.G. (1977) ‘The measurement of observer agreement for categorical data’, *Biometrics*, 33(1), pp. 159–174. doi: 10.2307/2529310.

LGFA (2024) ‘About us – Ladies Gaelic Football’. Available at: https://ladiesgaelic.ie.

Malone, S., McGuinness, A., Duggan, J.D., Murphy, A., Collins, K. and Burns, C. (2023), *Sport Sciences for Health*, 19(3), pp. 959–967.

Mangan, S., Collins, K., Burns, C. and O’Neill, C. (2022), *International Journal of Sports Science & Coaching*, 17(1), pp. 208–219.

Mangan, S., Ryan, M., Devenney, S., Shovlin, A., McGahan, J., *et al.* (2017) ‘The relationship between technical performance indicators and running performance in elite Gaelic football’, *Journal of Performance Analysis in Sport*, 17(5), pp. 706–720.

Martin, D. (2004) ‘Norms and trends in Gaelic football’, in *Proceedings of Performance Analysis of Sport IV*. Belfast: ‹Publisher/host›, pp. ‹pages›.

McColgan, A., Bradley, J., McKenna, O., Byrne, A. and Twomey, L. (2024) ‘Investigating successful kickout performance in senior inter-county Ladies Gaelic Football’, *International Journal of Performance Analysis in Sport*, 25(3), pp. 405–418. (Published online 21 October 2024). doi: 10.1080/24748668.2024.2416738

McDermott, L., Collins, K., Mangan, S. and Warne, J. (2021) ‘The influence of opposition quality on team performance indicators in elite Gaelic football’, *International Journal of Performance Analysis in Sport*, 21(5), pp. 780–789.

McGuckin, B., Bradley, J., Hughes, M., O’Donoghue, P. and Martin, D. (2020) ‘Determinants of successful possession in elite Gaelic football’, *International Journal of Performance Analysis in Sport*, 20(3), pp. 420–431.

McGuigan, K. and Collins, K. (2021) ‘Understanding the impact of pitch location on shot outcome in Gaelic football – where is the scoring zone?’, *International Journal of Performance Analysis in Sport*, 21(4), pp. 491–506.

Mernagh, P., O’Donoghue, P., Martin, D., *et al.* (2019) ‘Does experience influence Gaelic football performance?’, *International Journal of Performance Analysis in Sport*, 19(6), pp. 947–960.

O’Donoghue, P. (2010) *Research Methods for Sports Performance Analysis*. London: Routledge.

O’Gorman, J., Seevinck, J., Daly, D., *et al.* (2020) ‘Digital transformation in sport: A call for coordinated applied research’, *International Journal of Sports Science & Coaching*, 15(2), pp. 147–150.

O’Leary, D., O’Shea, P., Keane, P., *et al.* (2023) ‘Optimising kick-out strategies in Gaelic football’, *European Journal of Sport Science*, 23(8), pp. 1203–1214.

Ostrek, M. and Messerli, T. (2019) ‘DeepBall: Robust ball detection in broadcast football videos using hypercolumn features’