MATH Sport 2026

18TH AUSTRALASIAN CONFERENCE ON MATHEMATICS AND COMPUTERS IN SPORT

Hosted by UniSC and Bond University

Proceedings
Including

CONFERENCE SCHEDULE
& FEATURED SPEAKER BIOGRAPHIES

1-3 JULY 2026 | GOLD COAST, QUEENSLAND

UniSC | BOND UNIVERSITY

1

MATHSport 2026

18th Australasian Conference on Mathematics and Computers in Sport | 1-3 July 2026

FEATURED SPEAKER BIOGRAPHIES

Day One keynote and Day Two Women in Maths & Computers in Sport Forum

Featured sessions connect mathematical and analytical insight with lived experience at the highest levels of sport.



David Kendix

ICC statistician consultant; President, Association of Cricket Statisticians and Historians

Day One Keynote: *Turning theory into practice: three models that changed cricket*

David Kendix is an internationally recognised cricket statistician and a Fellow of the Institute of Actuaries. As statistician to the International Cricket Council and a long-serving member of its Cricket Committee, he identifies statistical trends, contributes to playing regulations and translates quantitative ideas into systems used across the international game. He peer-reviewed the Duckworth-Lewis-Stern method before its adoption and designed the ICC team rankings, which now cover more than 100 men's and women's teams. David has also served as an expert witness in a match-fixing trial, scored more than 100 internationals including 50 Tests, and held board and treasury roles at Middlesex County Cricket Club. Beyond cricket, he designed and maintains World Netball's team rankings and helped shape its world championship tournament structure.

CORE EXPERTISE	CRICKET IMPACT	LEADERSHIP
Model risk & rankings	ICC systems since 2002	ACS President since 2017



Lilliana Prucha

Australian Steelers wheelchair rugby athlete; Bond University data analytics student

Day Two Forum: *Her Game, Her Numbers: Tales from Elite Sport*

Lilliana Prucha is a rising member of the Australian Steelers wheelchair rugby program and one of a new generation of athletes combining elite performance with data-driven thinking. She made her senior Steelers debut at the 2025 World Challenge and was selected for Australia's 2026 Canada Cup campaign. After relocating from Perth to the Gold Coast, Lilliana joined Queensland's high-performance training environment and has continued to develop her tactical awareness, decision-making and defensive impact. She is the first wheelchair rugby athlete to receive Bond University's John Eales Rugby Excellence Scholarship and is completing a Bachelor of Data Analytics with a minor in Sports Analysis. Away from competition, she coaches junior wheelchair rugby and is interested in how technology and analytics can improve athlete performance and understanding.

TEAM	CLASSIFICATION	STUDY
Australian Steelers	2.0F	Data Analytics + Sport Analysis

Bio sources: MathSport Asia / ICC / ACS; Paralympics Australia and Bond University. Current to June 2026.

UniSC | BOND UNIVERSITY

FEATURED SPEAKER BIOGRAPHIES

Her Game, Her Numbers: Tales from Elite Sport | Thursday 2 July, 9:30 am

Featured sessions connect mathematical and analytical insight with lived experience at the highest levels of sport.



Zoe Hanna

Queensland Reds back-rower; Bond University Doctor of Physiotherapy student

Day Two Forum: *Her Game, Her Numbers: Tales from Elite Sport*

Zoe Hanna is a dynamic back-rower with the Queensland Reds and a key figure in Bond University's highly successful women's rugby program. In 2022 she became the first female recipient of Bond's prestigious John Eales Rugby Excellence Scholarship, helping open a new pathway for women in the sport. Zoe has been central to the Bull Sharks' premiership success and entered her second Super Rugby Women's season with the Reds in 2026 after returning from shoulder surgery. Her performances have also earned selection in Wallaroos training environments and Australia A representation. Alongside rugby, Zoe completed a Bachelor of Exercise and Sports Medicine and is now studying a Doctor of Physiotherapy, giving her a valuable athlete perspective on preparation, recovery, injury management and the practical use of performance information.

POSITION	TEAM	STUDY
Flanker / No. 8	Queensland Reds	Doctor of Physiotherapy



Laura Langman ONZM

Silver Ferns great former Sunshine Coast Lightning and New Zealand captain

Day Two Forum: *Her Game, Her Numbers: Tales from Elite Sport*

Laura Langman ONZM is one of the most influential and respected midcourters in world netball. Across a remarkable international career she earned 165 Test caps - the most in Silver Ferns history - and played centre, wing defence and wing attack with relentless work rate, speed and tactical intelligence. After returning to international netball in 2018, Laura captained New Zealand to the 2019 Netball World Cup title in Liverpool. Her career also included Commonwealth Games gold medals, domestic championships in New Zealand and Australia, and a leading role with the Sunshine Coast Lightning. Renowned for her professionalism, durability and ability to influence both attack and defence, Laura brings a rare athlete's view of how physical data, tactical analysis, communication and trusted analyst relationships can support sustained elite performance.

SILVER FERNS	MAJOR TITLE	HONOUR
165 International Test caps	2019 World Cup champion	ONZM for services to netball

Bio sources: Bond University and Queensland Reds; Silver Ferns and New Zealand Olympic Team. Current to June 2026.

UniSC | BOND UNIVERSITY

MATH Sport 2026

18th Australasian Conference on Mathematics and Computers in Sport | 1-3 July 2026

DAY ONE - WEDNESDAY 1 JULY 2026

Sofitel Gold Coast Broadbeach - Lobby and Sorrento Room, Level 3

25-minute talks: 20 minutes + 5 minutes Q&A. 15-minute talks: 12 minutes + 3 minutes Q&A. ★ Powerplay contribution.

The first page of each abstract or paper is shown in parentheses after each session title

09:00-13:35

TIME	MIN	AREA / PRESENTER	SESSION TITLE- Chair Steve Stern
09:00	20	**Lobby / Sorrento Room**	Registration
09:20	10	**Welcome** Steve Stern + Ant Bedford	Welcome, tipping competition, housekeeping, next conference and dinner sign-up
09:30	55	**Cricket 1** David Kendix	**Turning theory into practice: three models that changed cricket**
10:25	25	**Cricket 2** Marc West	Connecting the dots: a mathematical analysis of Bazball (5)
10:50	20	**MORNING TEA**	
11:10	25	**Cricket 3** Mingzhong Wang	BPARC: Bilaterally player-aware resource compensation for rain-interrupted limited-overs cricket matches (15)
11:35	25	**Racquet Sports** Yuki Takasaki	Massey ratings for match outcome prediction in table tennis: evidence of greater stability than the ITTF World Ranking (29)
12:00	25	**Racquet Sports** Ken Louie	Scheduling interclub squash leagues in New Zealand to improve player engagement (38)
12:25	45	**LUNCH – Session Chair Darren O'Shaughnessy**	
13:10	25	**AFL** Hamish Murray	List management optimisation in the Australian Football League (45)
13:35	25	**AFL** Steven Azzopardi	Comparative analysis of defensive pressure in men's and women's Australian football (46)

14:00 ONWARDS

TIME	MIN	AREA / PRESENTER	SESSION TITLE
14:00	25	**AFL** Dave Matteo	Approaches to measuring and modelling player development in Australian rules football (47)
14:25	25	**AFL** Paul Hardwicke + Ant Bedford	Save our scores: offsetting AFL strength-of-schedule imbalance with points recalibration (48)
14:50	20	**AFTERNOON TEA – POWERPLAY – Chair Ant Bedford**	
15:10	25	**Hydration Sensing** Alex Lap Cheung Liu	Towards hydration sensing: a dimension-reduction framework for impedance spectroscopy data (57)
15:35	25	**Swimming** Mackenzie Parker	A data-informed framework to classify and analyse pacing profiles in competitive swimming (58)
16:00	25	**Payments** Mana Mirzaei	Mathematical modeling of adaptive payment rail switching in high-demand sports events (59)
16:25	25 ★	**Invasion / Rugby / Quantum** Sam Forbes	Regrets, I've had a few. (69, 77, 85, 93, 99)
16:50		**Suggested dining at The Star: a table will be booked for 20 guests at Uncle Su (Asian). Pay as you go; confirm at registration.**	

UniSC | BOND UNIVERSITY

CONNECTING THE DOTS: A MATHEMATICAL ANALYSIS OF BAZBALL

Marc West, Bernard Kachoyan [a,b]
[a] *University of New South Wales*
[b] *Corresponding author: b.kachoyan@unsw.edu.au*

Abstract

England's men's Test batting resurgence under "BazBall" is widely portrayed as an aggressive, boundary-driven shift in philosophy. Using ball-by-ball data from the 46 Tests following England's coaching change and a matched pre-period, we compare England, their opponents, and a contemporaneous control group. England's increased scoring rate is driven primarily by a substantial reduction in dot balls rather than by an increase in fours and sixes; the distribution of scoring shots shows little change. Across world cricket there is a small decline in dot-ball percentage over the same window consistent with historical variation, but England's dot-ball reduction exceeds ten percentage points and is far outside normal drift. England changed *when* they score, more than *how* they score.

Using a survival-analysis framework, we demonstrate that the rise in England's batting average exceeds what would be expected from faster scoring alone. Although dismissal probability per ball rises, it increases too little to offset the higher expected score generated by more frequent scoring opportunities. Decomposing the change in average shows that most of the increase is due to reduced dismissal risk per scoring opportunity, suggesting an improvement in underlying batting quality, with only a minor contribution from boundary frequency. We observe similar but smaller effects in England's opponents while the control group shows little corresponding change, suggesting the phenomenon is specific to England matches rather than a global trend. The shift also increases decisiveness at match level: draws become rare, England's wins rise, and losses remain unchanged.

Finally, we place these results in historical context: despite major changes in the game, Test cricket exhibits a self-regulating scoring/risk equilibrium, and how it adapts to BazBall remains an open question.

Keywords: cricket, survival modelling, batting, BazBall

1. INTRODUCTION

"BazBall", named for England men's cricket coach Brendon "Baz" McCullum, refers to the playing philosophy of the England men's cricket team under McCullum and captain Ben Stokes [1, 2]. The period is not only notable for England's markedly higher scoring rate [3, 4], but also the extraordinary volume of narrative associated with it. We won't relitigate this here, but it's worth cutting through the noise. While data analysis cannot adjudicate the broader rhetoric, it can say a lot about on-field performance. Did England's batting performance improve under BazBall? Did it translate into more team success? And where does it sit within the history of Test cricket? The 46-Test BazBall window provides a large enough sample to address these issues directly without contributing to the hot-take cycle.

2. SURVIVAL MODEL

The survival model of Kachoyan and West was used to model batting as a stochastic survival process with discrete scoring outcomes [5]. The probability of scoring (surviving) more than j runs is:

$$S_j = P(\text{score} > j) \tag{1}$$

Therefore, the probability of scoring more than zero is:

$$S_0 = P(\text{score} > 0) \tag{2}$$

It was shown previously that S_0 is an initial condition that should be treated as a separate, empirically derived, parameter in cricket batting survival modelling [6]. That is, the probability of scoring zero is a special case as players are vulnerable early in their innings. It has been shown separately that most batters play with less than

half their potential batting ability at the start of their innings [7]. This can be empirically estimated from innings data:

$$S_0 = 1 - \frac{N_0}{N} \tag{3}$$

where:
N_0= number of innings scoring 0 (out or not out)
N= total innings.

Note that we are using S_0 to model behaviour at 0. Therefore, we define μ as the dismissal risk per scoring opportunity when on scores greater than 0.

$$\mu = \frac{D - D_0}{B - B_*} \tag{4}$$

where:
D= total dismissals
D_0 = total dismissals on 0 (ducks)
B = total balls faced
B_* = total dot balls

Here we assume μ is constant, however the full model allows for varying μ at every step [5]. The implications of this are discussed further in the results.

The probability of scoring k runs off a scoring shot is

$$P_k = P(\text{ score } k \mid \text{ scoring shot }) \tag{5}$$

for $k \in \{1,2,3,4,5,6\}$, with

$$\sum_{k=1}^{6} P_k = 1 \tag{6}$$

We ignore scores of more than 6 off a ball as their occurrence is vanishingly small.[1] The parameter κ is the expected number of runs conditional on scoring and is therefore determined entirely by the distribution of scoring-shot outcomes P_k.

$$\kappa = \sum_{k=1}^{6} k\, P_k \tag{7}$$

The first-order expansion of the survival expression in Kachoyan and West yields the expected batting average. Because observed parameter changes are small, higher-order terms are negligible; see [5] for the full derivation. It is this first order equation that forms the basis of the analysis in this paper.

$$\alpha \approx \frac{S_0\, \kappa}{\mu} \tag{8}$$

[1] There was a single 7 observed in the pre and Bazball periods: Will Young for New Zealand vs. Bangladesh, Christchurch 2022, where he edged to slip, was dropped, ran 3 and then had 4 overthrows. The highest score off a single ball we could find in the cricsheet dataset, which goes back to 2001, was Andrew Symonds hitting an 8 (4 all run then 4 overthrows) during the Australia vs. New Zealand Test Match at the Gabba in 2008. Quirks like this can be found throughout Test history if you look hard enough. We found that Michael Clarke has the distinction of having two 7s against his name.

3. DATA

Ball-by-ball data are sourced from Cricsheet[2] [8]. The BazBall period consists of the 46 England Tests from June 2022 to January 2026. The comparison period is the preceding 46 Tests beginning in August 2018. A contemporaneous control cohort includes all non-England Tests played during these same windows (192 Pre-BazBall, 202 BazBall). Table 1 shows the raw number of balls that were hit for 0 to 6 runs, and Table 2 shows the summary statistics for the periods under investigation. All calculations are performed on pooled balls, not per Test.

Table 1: Raw scoring ball distribution counts

	England		Opposition		Control	
Runs	PRE	BAZ	PRE	BAZ	PRE	BAZ
0	33201	23025	33339	32701	136198	124599
1	6411	7357	6174	7117	25545	28934
2	1536	1669	1528	1501	6532	6044
3	354	342	374	437	1683	1364
4	2413	2895	2452	2980	10147	10126
5	5	3	8	6	29	24
6	175	298	174	303	755	1054

Table 2: Summary statistics

	England		Opposition		Control	
	PRE	BAZ	PRE	BAZ	PRE	BAZ
Dot ball rate	75.3%	64.7%	75.7%	72.6%	75.3%	72.4%
P_1	58.9%	58.6%	57.7%	57.7%	57.2%	60.9%
P_2	14.1%	13.3%	14.3%	12.2%	14.6%	12.7%
P_3	3.3%	2.7%	3.5%	3.5%	3.8%	2.9%
P_4	22.2%	23.0%	22.9%	24.1%	22.7%	21.3%
P_5	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%
P_6	1.6%	2.4%	1.6%	2.5%	1.7%	2.2%
κ	1.95	2.00	1.98	2.04	1.99	1.94
Average	26.82	33.65	27.84	30.36	29.37	27.70
Strike Rate	48.24	70.54	48.23	55.94	49.16	53.48

4. DOT BALLS AND SCORING SHOTS

Table 2 shows that England's strike-rate increase is overwhelmingly explained by a reduction in dot balls rather than an increase in boundaries; during BazBall, England scored off every third ball, as opposed to every fourth. The increase in scoring shot percentage (25% to 35%) is several times larger than global drift (25% to 28%), demonstrating that BazBall's defining statistical feature is dot-ball suppression rather than boundary inflation. This is shown in Figure 1.

Notably, the distribution of scoring outcomes conditional on scoring (P_k) changes little across eras, except for a small increase in 6s. The structure of scoring shots is effectively constant, with κ remaining close to 2 in all scenarios. There is a small increase in 6s in line with historical trends. The implication is that BazBall primarily changes *whether* runs are taken rather than *how* they are scored. This is shown in Figure 2.

A possible interpretation for this change is to assume that some dot-balls are a choice by the batter, not strictly due to those balls being "too good" to score from. If we assume this to be the case, we can see that the additional non-dot balls created have a run-outcome mix that is much the same as the baseline scoring distribution. This is scoring more often without altering shot selection, as opposed to say a boundary-driven revolution or an increase in opportunistic singles (two common explanations for an increased run rate). The model has some merit – Test batters spend time in the middle wearing down their opponents and may well choose to leave balls as part of fatiguing bowlers. However, it is not completely satisfactory – would batters be

[2] Cricsheet is an open-source ball-by-ball database that is complete for the time periods under-examination, except for Afghanistan Tests, which are excluded. Wides and no-balls are included in the ball count, but the number of runs scored is only runs off the bat. Hence a dot-ball as described in this paper is 0 runs off the bat; it could be a wide or a no-ball, however the effects are expected to be negligible. Data is available on request to the authors.

deliberately leaving "bad balls" they could otherwise have hit for four or six? The small uptick in boundaries could take this into account, but the effect is small.

A separate plausible contributor is match ecology: if the game tempo influences field settings, bowling plans, and fatigue, more scoring opportunities for both sides may arise, a possibility we return to in the discussion. Again, the model is plausible but the lack of change in the scoring shot distribution is curious. It is likely that this result is a combination of both ideas.

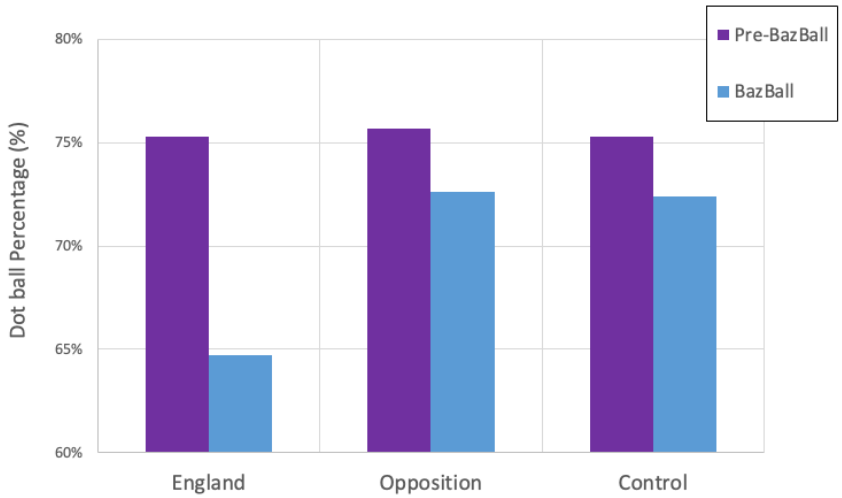


Figure 1: Dot-ball percentage for England, their opposition and the control group across the pre-BazBall and BazBall eras.

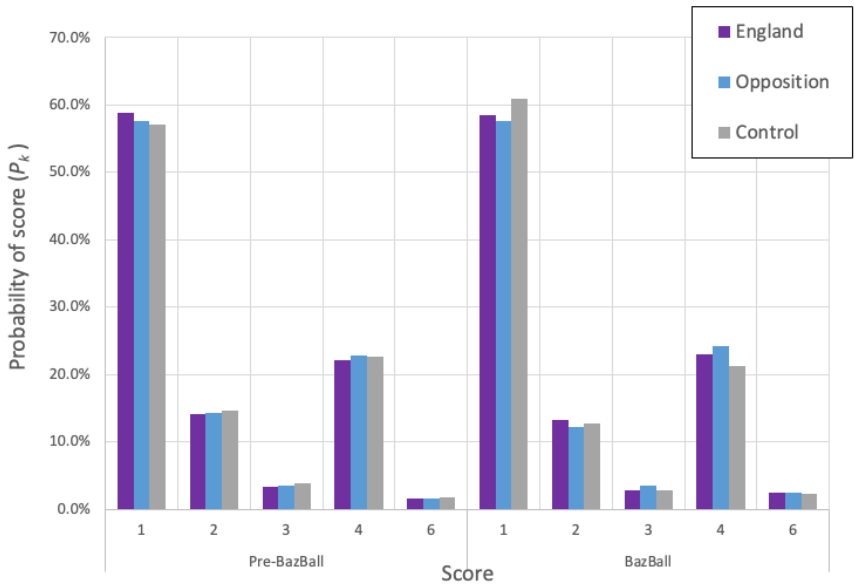


Figure 2: Conditional scoring outcomes (P_k) for England, their opposition and the control group, across the pre-BazBall and BazBall eras. 5s have been removed for clarity (they are negligible as per Table 1).

5. BATTING IMPROVEMENTS

The next step in analysing BazBall batting is to examine whether, along with the increased run rate, more runs are being scored. To do this, we use the survival model defined earlier in Equation (8). We can then perform a first-order log-linear decomposition.

$$\Delta \log \alpha \approx \Delta \log S_0 + \Delta \log \kappa - \Delta \log \mu \tag{9}$$

$$\Delta \alpha \approx \alpha_{\text{pre}} (\Delta \log S_0 + \Delta \log \kappa - \Delta \log \mu). \tag{10}$$

We note the contribution of S_0 to the change in batting average as $\alpha_{\text{pre}}(\Delta \log S_0)$ and call it the "S_0 *Effect*" in Table 4. Similarly, for κ and μ, noting that the μ *Effect* has the reverse sign; an increase in risk will decrease the average. The residual is the difference between the actual average change and the sum of the various effects[3]. Table 3 shows the parameters being used and Table 4 shows the results.

Table 3: Survival modelling parameters

Cohort	Era	μ	S_0	κ
England	PRE	0.0623	0.847	1.95
England	BAZ	0.0536	0.896	2.00
Opp	PRE	0.0615	0.860	1.98
Opp	BAZ	0.0595	0.873	2.04
Control	PRE	0.0579	0.848	1.99
Control	BAZ	0.0611	0.869	1.94

Table 4: Decomposition of average change

Cohort	Ave PRE	Ave BAZ	S_0 Effect	κ Effect	μ Effect	Residual
England	26.82	33.65	+1.52	+0.62	**+4.02**	+0.66
Opposition	27.84	30.36	+0.43	+0.80	+0.95	+0.33
Control	29.37	27.70	+0.70	−0.80	−1.55	−0.01

The first-order decomposition shows that the increase in England's batting average is driven primarily by a reduction in dismissal risk per scoring opportunity, captured by the parameter μ. We already see that μ falls from 0.062 to 0.054, suggesting a smaller risk per scoring shot. Now we can see that of the total rise of roughly seven runs per wicket, approximately four runs are attributable to this improved conversion of scoring opportunities into continued survival. Improvements in early survival probability (S_0) account for a smaller share of the increase, while changes in the scoring-shot value parameter κ contribute only modestly as κ does not significantly change. The dominant factor in England's batting improvement is that batters are dismissed less frequently relative to how often they score, indicating a shift in underlying batting effectiveness.

England's opponents exhibit the same directional pattern but with smaller magnitudes, suggesting that some portion of the effect reflects the match environment. By contrast, the control group shows declining averages driven by declining effectiveness and shot value despite modest gains in early survival. This suggests that the batting improvement is not part of a global trend but a shift occurring specifically in matches involving England.

Residual differences between the linearised model breakdown and observed averages are modest second order effects of linearisation; their small size suggests that the survival decomposition captures the dominant structure of batting outcomes well.

If we simply use Equation (4) rather than combining the results of linearisation, the predicted averages tend to slightly underestimate the averages, while the predicted average differences match well. Table 5 shows that the underprediction is ~0.1–0.4 across cohorts and is broadly stable across windows. The under-estimate is entirely consistent with a well-known Jensen-convexity effect in survival mixtures and is due to treating the team of 11 players as homogenous as opposed to being comprised of different abilities (a mixture / heterogeneity bias) [9, 10]. Additional contributors include changes in μ throughout a batter's innings [11] and tail effects [6].

Table 5: Comparison of actual and model averages

	Actual			Model		
Cohort	Ave PRE	Ave BAZ	Δ Ave	Ave PRE	Ave BAZ	Δ Ave
England	26.82	33.65	**+6.83**	**26.52**	33.38	+6.85
Opposition	27.84	30.36	**+2.52**	**27.72**	29.99	+2.26
Control	29.37	27.70	**−1.67**	**29.14**	27.54	-1.60

[3] That is, the residual is the linearisation remainder.

6. DO ENGLAND BAT LONGER THAN EXPECTED?

To test whether England's higher batting average is explained purely by faster scoring, we construct a counterfactual in which scoring behaviour changes under BazBall but dismissal risk per scoring shot (μ) above zero does not. We then examine whether batters are facing more balls than you might expect. Noting again that the model treats scores at zero separately, this modelling is for scores above 0. To determine the counter-factual, we need to know how many balls those who scored zero (out and not out) faced. We define this as B_0. Thankfully, we have this data[4].

We define dismissal hazard per ball (h_b) as:

$$h_b = \frac{D - D_0}{B - B_0} \tag{11}$$

where
B_0= total balls faced by those who scored 0 (out and not out)

We define the probability that a run is scored off a ball, excluding 0-innings balls, as *s*:

$$s = \frac{B - B_*}{B - B_0} \tag{12}$$

Using the subscript *cf* to denote *counterfactual*, since balls per dismissal (BPD) $= \frac{1}{h_b}$, holding μ fixed at pre-BazBall levels gives:

$$h_{b,\mathrm{cf}} = \mu_{\mathrm{pre}}\, s_{\mathrm{baz}} \tag{13}$$

Table 6 shows the counterfactual results.

Table 6: Counterfactual results table

Cohort	BPD BAZ	BPD CF	Extra Balls
England	52.30	45.01	+7.29
Opposition	60.67	58.62	+2.04
Control	61.73	58.54	-3.18

Table 6 shows that England batters survive **~7 balls longer per dismissal than expected** given their scoring rate, given they score more than 0. England's opponents show a smaller but similar effect (~+2 balls), while the control cohort shows an opposite pattern (~-3 balls). It indicates that the BazBall effect is specific to matches involving England rather than a global trend.

The BazBall effect is especially pronounced near zero. We saw earlier that England's probability of surviving beyond zero (S_0) rose in BazBall, and here we can also show that the total number of balls consumed by scoreless innings falls. This is consistent with more proactive "getting off the mark" behaviour. With a higher scoring rate, fewer innings remain trapped at zero long enough for dismissal to occur, and fewer innings end scoreless at all. At the same time, scoreless innings that do occur end faster on average, suggesting a reduced emphasis on hanging around on 0. The combination of higher tempo and improved S_0 indicates that BazBall changes not only middle and latter over scoring but also early-innings dynamics.

[4] We have sourced some data from cricinfo Statsguru – there may be differences with cricsheet, but they are tiny.

6. HISTORICAL CONTEXT

The survival identity used to decompose BazBall's batting improvement – Equation (8) – also offers a useful lens for interpreting Test cricket historically. In this framework, average is expressed in terms of early survival beyond zero (S_0), expected runs per scoring shot (κ), and dismissal risk per scoring opportunity (μ). We restate the relationship here for easy reference:

$$\alpha \approx \frac{S_0\,\kappa}{\mu} \tag{14}$$

This equation makes explicit a form of self-regulation: if scoring-shot value increases, average will rise unless dismissal risk increases commensurately or early survival falls. This broader idea of a batting speed/survival trade-off has recently been examined by Rizvi [12], who relates run-scoring rate to innings half-life in pooled international player data through an approximate power-law relationship. Our focus here is different: rather than estimating a player-level frontier across formats, we consider the aggregate historical evolution of Test batting and the striking stability of batting average under changing scoring conditions as a way of putting BazBall into historical context. Using data from cricinfo [13], Figure 3 shows that over long periods of Test history the overall Test batting average is remarkably stable, not moving much beyond ~30 and bounded within a narrow band. The chart groups runs in 5-year bins and plots back to 1910, because before this time, Tests were relatively uncommon (<24 per bin). Hence average is defined by total runs divided by total dismissals within the bin. The added flat line is a least-squares fit to the data, showing that historically, the average trend is flat (slope = 0.002, R^2 = 0.00). The empty bins are World Wars One and Two. The amount of data in each bin reduces as we go back in time due to less Tests being played, so therefore we see more variation.

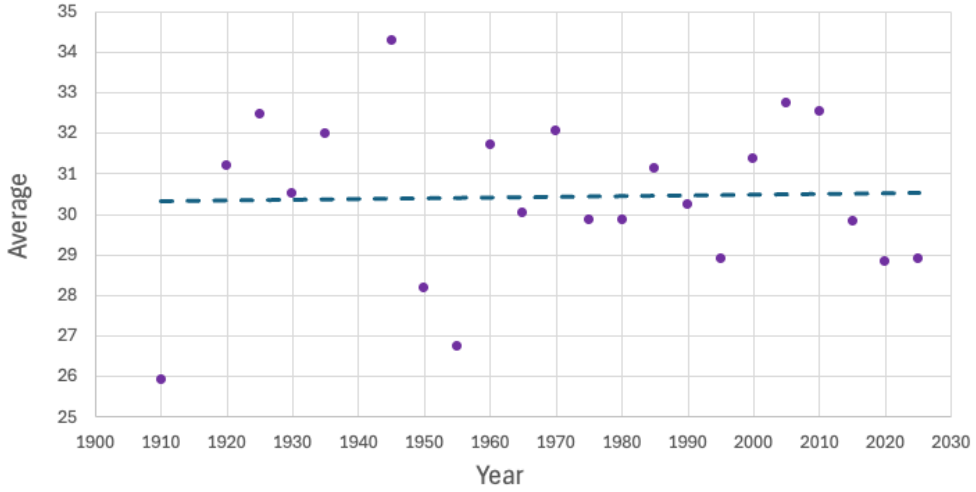


Figure 3: Historical Test batting average (5-year bins, since 1910)

It is not possible to plot κ and μ for the whole timeline without ball-by-ball data, or at least better summary data than is available. However, boundaries are reported for many Tests, and the percentage of runs from boundaries is an observable indicator of the scoring-shot environment. Figure 4 shows the percentage of total runs from 4s and 6s (where boundaries are reported). 4s climbed to the 1960s and are roughly flat since, while 6s have increased continually. 6s are reported on the secondary axis.

μ depends on scoring-frequency information that is typically unavailable pre-2000. Nonetheless, a proxy for μ is hazard per ball, and an intuitive way of viewing this is through its reciprocal – balls per dismissal. We can recreate this from scorecards with total balls reported; data pre-1960 is extremely patchy (for example, there is only one Test during 1950-1954 with total balls reported). Figure 5 shows results for periods with more than 5 Tests (per 5-year bin). We see an almost linear decrease in the number of balls faced per dismissal[5] (slope = - 0.36, R^2 = 0.83).

[5] Note that unlike our earlier analysis of Bazball, for this calculation we are looking at all scores, not just scores > 0. This is for simplicity: here we are showing proxies as we do not have enough data to construct μ.

S_0 can be constructed from available data and suggests a gentle reduction over time, although the data is noisy despite the 5-year bins and the estimated slope is sensitive to the choice of start period. Hence, we do not consider it a major driver in the system compared to the other factors.

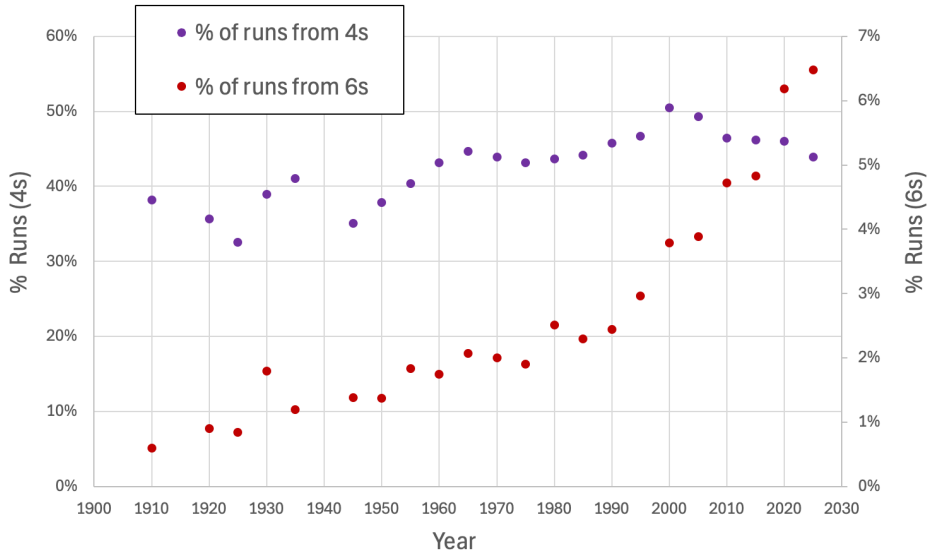


Figure 4: Historical percentage of runs from 4s and 6s (5-year bins, since 1910)

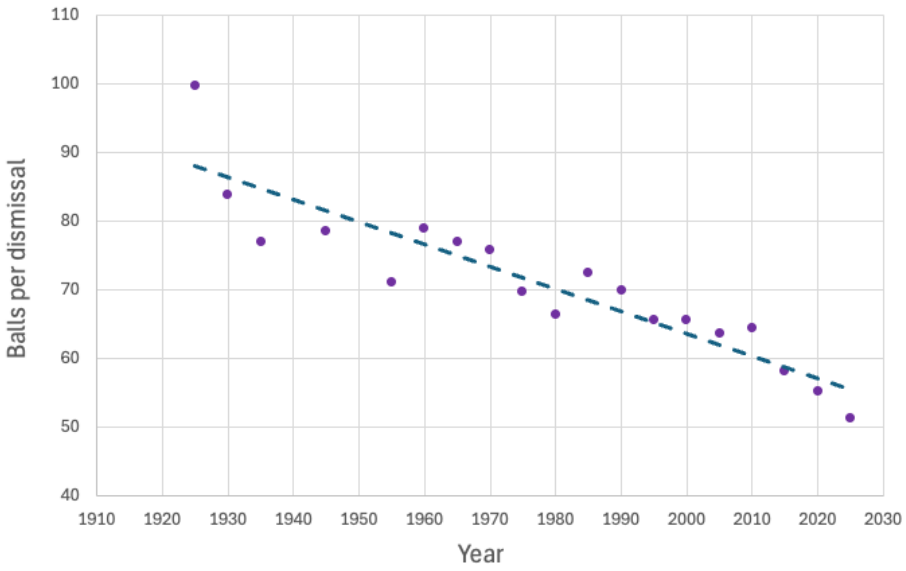


Figure 5: Historical balls per dismissal (5-year bins, only periods with >5 Tests reported, since 1910)

A natural interpretation of these results is that Test cricket behaves like a coupled adaptive system. Improvements on the batting side – for example, better bats, covered pitches and smaller grounds producing a higher percentage of boundaries – do not occur in isolation. Bowling and fielding also evolve through tactical responses and fielding proficiency. This is akin to an arms race or a predator-prey dynamic: when batters become more effective at extracting runs, bowlers and captains adapt to restore control, and vice versa. The consequence is not stasis but stability in the *aggregate* outcome produced by continual co-evolution.

Distinguishing these mechanisms is a natural direction for future work, and more generally the identity linking S_0, κ, and μ – equation (14) – suggests a broader program analysing Test cricket as a dynamically self-regulating system shaped by co-evolution between batting and bowling. Future work could also extend the single-μ model here into one that considers the heterogeneity of a team of 11 batters [14], building on the work of Kachoyan and West in [15].

7. THE PLACE OF BAZBALL

We now examine where BazBall sits within this adaptive system. The modern game is shaped by a desire for more results and increasingly played in the shadow of limited overs as cricket moves into the entertainment business [16]. Modern incentives like the World Test Championship mean there are no dead rubbers; teams are less willing to play for draws and more willing to trade safety for tempo, and consequently, Tests are getting shorter [17]. This shifts the system toward higher run rates and higher dismissal risk per ball. As discussed, it is striking that long-run batting average remains broadly stable despite these external pressures. Within this context, BazBall appears as a deviation from the risk/tempo coevolution. England's tempo increase is driven primarily by a sharp reduction in dot balls, while P_k and κ are largely unchanged. In the decomposition, average rises because dismissal risk does not increase enough to offset the higher expected accumulation implied by faster scoring. In survival terms, England shift to a higher-tempo operating point without paying the full historical hazard penalty, consistent with improved underlying batting performance.

Our earlier ecological framing helps interpret the observation that England's opponents show similar, if smaller, improvements. If a match's tempo is jointly determined, then one side's approach can change the equilibrium conditions faced by both teams. BazBall may therefore operate partly as a match-level perturbation: England increase the frequency of scoring opportunities (reducing dot balls), opponents respond with different containment and wicket-taking trade-offs, and the resulting environment shifts the scoring patterns of both sides. This does not explain away England's improvement – opponents' gains are smaller – but it does suggest BazBall is not purely an isolated, team-internal "philosophy change"; it also alters the ecology of the contest.

Team results align with this view. Table 1Table 7 shows that across the matched 46-Test windows, England's win count rises from 19 to 26, while losses remain unchanged at 18 and draws fall from 9 to 2. This pattern is consistent with BazBall increasing decisiveness: higher tempo reduces the time constraint that historically produces draws, increasing the probability of a result.

Table 7: Match results

	Tests	Wins	Losses	Draws	Win %	Loss %	Draw %
Pre-BazBall	46	19	18	9	41%	39%	20%
BazBall	46	26	18	2	57%	39%	4%

8. FINAL REMARKS

Cricket is an eccentric sport, historically and culturally. A middle-class pursuit exported across the world through colonialism, it still holds onto quaint Englishness like the tea-break; the "Gentlemen vs. Players" ethos of the past is not lost, even with many countries using cricket to express themselves in the post-colonial era [18]. To many, the "spirit of cricket" operates above the written laws of the game [19], even if this is probably a fiction. Nonetheless, cricket is never just about winning; it is also about the manner of play and the aesthetics. As McKenna noted, there is "an irresolvable gap between the objective goal of winning the game and the subjective quality of cricketing excellence" [20].

It is within this context that we place BazBall. Objectively, the story is straightforward: averages and scoring rates rise and England wins more often. But subjectively, it is far less clear. Even the authors of this paper cannot agree on whether England "bat better" under BazBall. The era gives you Harry Brook rewriting early-career records and Jamie Smith being bounced out by Marnus Labuschagne. It is both beautiful and ugly. But should that actually matter, does it have to look "like cricket?" And how will the grand auld game of Test Cricket adapt? As Edgar rather beautifully states, "players find new ways of playing by the old rules, and thereby they find new ways to answer the challenges those rules pose - new ways to explore the potential of being human" [21]. One thing we do know is that copious words will be written about it.[6]

[6] We note the irony of writing about it here.

9. REFERENCES

[1] Switch Hit Podcast, “Let's play... Baz-Ball!,” ESPN, 2022.

[2] S. Bowman, Safe to Great: The New Psychology of Leadership, Vancouver: Figure 1 Publishing Inc., 2023.

[3] T. Newans and C. Drovandi, “Will Bazball finally knock its critics for six?,” *Significance,* vol. 21, no. 1, pp. 32-35, 2024.

[4] N. Hoult and L. Booth, BazBall: The Inside Story of a Test Cricket Revolution, London: Bloomsbury, 2023.

[5] B. Kachoyan and M. West, “Batting risk-reward tradeoffs in cricket,” *Parabola,* vol. 61, no. 2, 2025.

[6] B. Kachoyan and M. West, “Deriving an exact batting survival function in cricket,” in *14th Australasian Conference on Mathematics and Computers in Sport (ANZIAM Mathsport 2018)*, Sunshine Coast, 2018.

[7] O. G. Stevenson and B. J. Brewer, “Bayesian survival analysis of batsmen in Test cricket,” *Journal of Quantitative Analysis in Sports,* vol. 13, no. 1, 2017.

[8] S. Rushe, “Cricsheet,” [Online]. Available: https://cricsheet.org/. [Accessed 19 February 2026].

[9] M. Xu, “Negative dependence in frailty models,” *Journal of Statistical Planning and Inference,* vol. 138, no. 5, pp. 1433-1441, 2008.

[10] T. A. Balan and H. Putter, “A tutorial on frailty models,” *Statistical Methods in Medical Research,* vol. 29, no. 11, 2020.

[11] B. Kachoyan and M. West, “Cricket as life and death,” in *13th Australasian Conference on Mathematics and Computers in Sport (ANZIAM MathSport 2016)*, Melbourne, 2016.

[12] M. S. Rizvi, “Speed-performance tradeoff and its effect in the game of cricket: Predictive modeling from a statistical mechanics perspective,” *Physical Review E,* vol. 112, no. 6, p. 064121, 2025.

[13] “Statsguru,” 2026. [Online]. Available: https://stats.espncricinfo.com/ci/engine/stats/index.html. [Accessed 3 March 2026].

[14] P. Hougaard, “Frailty models for survival data,” *Lifetime Data Analysis,,* vol. 1, pp. 255-273, 1995.

[15] B. Kachoyan and M. West, “Predicting the expected batting partnerships in cricket,” in *15th Australasian Conference on Mathematics and Computers in Sport (ANZIAM Mathsport 2020)*, Virtual Conference, 2020.

[16] S. Duncan and T. Heenan, “Big bash as mass deception: understanding the dumbing down of cricket,” *Sport, Ethics and Philosophy,* pp. 1-16, 2025.

[17] A. Narayanan, “The case of the shrinking Test match,” ESPN, 2024.

[18] S. Szymanski and T. Wigmore, Crickonomics: The Anatomy of Modern Cricket, Bloomsbury Publishing, 2022.

[19] S. Chopra and D. Coady, “Not Cricket,” *Sport in Society,* vol. 10, no. 5, pp. 729-743, 2007.

[20] J. McKenna, “Introduction. At the Boundaries of Cricket,” *Sport in Society,* vol. 10, no. 5, 2007.

[21] A. Edgar, “The Death of Test Cricket,” *Sport, Ethics and Philosophy,* vol. 14, no. 2, pp. 127-128, 2020.

BPARC: BILATERALLY PLAYER-AWARE RESOURCE COMPENSATION FOR RAIN-INTERRUPTED LIMITED-OVERS CRICKET MATCHES

Abdul Mateen[a], Hassan Ahmed[a], Shahrzad Saremi[b], Rania Shibl[c], Mansoureh Mirzaei[b,*], Anthony Bedford[b], Erica Mealy[b], Mingzhong Wang[b]

[a]*Department of Computer Science, National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan*

[b]*School of Science, Technology and Engineering, University of the Sunshine Coast, Brisbane, Australia*

[c]*Faculty of Science and Engineering, Southern Cross University, Gold Coast, Australia*

[*]*Corresponding author: mmirzaei@usc.edu.au*

Abstract

In cricket, rain interruptions in limited-overs matches force officials to revise the batting side's target. The current standard, Duckworth-Lewis Stern (DLS), adjusts targets based on remaining overs and wickets. However, although DLS accounts for overs remaining and wickets in hand through wicket-dependent resource curves, it does not distinguish between the identities or measured qualities of the specific players involved. In practice, top-order batters outscore tail-enders and are more reliable, and elite bowlers are more threatening and outperform part-timers. DLS therefore does not account for which specific batters or bowlers lose opportunities because of rain interruptions when determining whether a revised target is fair. Player-aware Resource Compensation in Interrupted Cricket Matches (PRCICM) [4] partially addressed this by weighting individual batter quality into the proposed model. It improves on DLS when strong or weak batting lineups are rain-affected. However, the model sees only one side of the game: the bowling side receives no quality adjustment and leaves a structural fairness gap. To overcome this limitation, this study introduces a technique, namely Bilaterally Player-Aware Resource Compensation (BPARC), which accounts for player quality on both sides simultaneously. This technique lowers the target where rain eliminates an elite bowler's remaining overs and vice versa for a weak bowler. A parallel mechanism governs batting quality on the chasing side. BPARC reduces the target exactly the same as in DLS when player quality on both sides is average to ensure a fair outcome. A key contribution of this study is the development of the Bowler Quality Index (BQI). Utilizing Multiple Linear Regression, the BQI synthesizes seven key performance indicators: economy rate, wicket rate, dot-ball percentage, strike rate, powerplay economy, death-over economy, and consistency into a single metric which captures both efficiency and consistency. This data is derived from ball-by-ball stats of 8,157 ODI and T20I matches (2010-2024). Simulation results show that on 412 DLS-affected matches, BPARC outperforms both baselines against an independent fairness benchmark (MAE $= 14.3$ runs vs. 18.6 for PRCICM and 22.1 for DLS). Whereas, win margins in successful chases tighten significantly ($p = 0.021$), and BPARC diverges from DLS by more than three runs in 67.2% of matches, confirming that player identity materially shapes what a fair revised target should be.

Keywords: cricket analytics, Duckworth-Lewis Stern (DLS), rain interruption, target revision, Bowler Quality Index, resource compensation, limited-overs cricket.

1. INTRODUCTION

With an estimated global audience of 2.5 billion people, cricket is one of the most widely followed sports in the world and has a significant professional tournament calendar having bilateral series, regional tournaments, and major events such as the ICC Cricket World Cups [1]. In limited-overs formats: One-Day Internationals (ODIs, 50 overs per side) and Twenty20 Internationals (T20Is, 20 overs per side), each team bats for a fixed number of overs and the side that scores more runs wins. Rain interruption is a fundamental problem in limited-overs cricket. As two match-playing sides do not face the same quota of matches. When overs are lost, one team has batted through the conditions, the other may never, and comparing raw or estimated run totals under those circumstances tells us very little about who actually performed better. Over the years, rain-affected decisions in important matches have even knocked teams out of ICC tournaments at the group stage, flipped bilateral series results, and left behind controversy. Players, coaches, fans, and data analysts question whether the revised targets were fair [6]. It not only impacts the results of matches, but the other stakes are very high: as they impact player rankings, prize money changes, and alter national records in ways that do not reflect players' genuine performance.

DLS, which is currently standard and adopted by the ICC in 1997 [2, 3]. Its logic is relatively simple: the batting resources available to the chasing side depend on two things: 1) overs remaining and 2) wickets in hand. And,

when an interruption removes some of those resources, the target shrinks proportionally. The framework has been revised twice since its introduction: once in 2003 and again in 2014. Steven Stern reworked resource curves to align them in line with the faster scoring that had come to define modern cricket [3]. DLS is straightforward to apply and is widely accepted. However, DLS remains intentionally player-agnostic: while it differentiates batting resources through overs remaining and wickets in hand, it does not distinguish between the identities or historical quality of the specific players involved. A batting lineup where the best five batters have not yet faced a ball is treated identically to another lineup with the same wickets and overs state, even if the remaining personnel are substantially weaker. Similarly, a bowling attack whose leading wicket-taker has three overs remaining is treated the same as one where only part-time bowlers have quota left. In each case, DLS gives the same revised target.

Research in cricket analytics has produced compelling evidence over the past two decades that player identity carries significant predictive information [10, 12, 15, 16, 17, 18]. Player-specific batting and bowling ratings, machine learning (ML) models trained on ball-by-ball data, and in-play win-probability tools all demonstrate that knowing *who* is batting or bowling materially changes the expected match outcome. Yet, despite this body of evidence, no target-revision method had incorporated individual player quality until Zia et al. [4] introduced PRCICM in 2022. PRCICM assigns each batter a Batter Resource Weight (BRW)—a relative quality score derived from historical performance data—and adjusts the DLS target by the collective BRW of the batters still to come at interruption. This was a meaningful step. However, PRCICM addresses only the batting side. The bowling side is left entirely unmodelled, meaning that an interruption which removes an elite fast bowler from the remaining overs is treated no differently from one that removes a tail-end part-timer.

This study introduces a framework named BPARC to fulfill the identified structural gap. The first target-revision framework to account for player quality on both sides of the match simultaneously. Here, the word "bilaterally" indicates that the adjustment responds to the quality of both the batting resources lost to the chasing side and the bowling resources lost to the fielding side. The main novel contribution of this research is a BQI, a composite player rating built from seven phase-specific features via Multiple Linear Regression. Our specific contributions are as follows.

- We develop the *BQI*, a composite measure of individual bowling strength derived from seven phase-aware features—economy rate, wicket rate, dot-ball percentage, bowling strike rate, powerplay economy, death-over economy, and over-to-over consistency—each capturing a distinct dimension of bowling contribution that career-average metrics cannot fully differentiate.
- We introduce the *BPARC formula*, a calibrated additive extension of the DLS target equation that integrates BQI alongside an extended version of the BRW from PRCICM. When $\psi = 1$ (average batting remaining) and $\phi = 1$ (average bowling remaining), BPARC reduces exactly to DLS, preserving backward compatibility with the official method.
- We evaluate BPARC against DLS and PRCICM on 412 real DLS-affected international matches using a four-metric framework designed to avoid the circular comparison that has affected all prior published target-revision evaluations—namely, the fact that a chasing team bats toward the official DLS target, making the second-innings score partially dependent on DLS rather than an independent measure of fairness.

The remainder of the paper is organized as follows. Section 2 surveys the related literature on target revision, player performance metrics, cricket analytics, and match outcome modelling, and identifies the four specific gaps that BPARC addresses. Section 4 describes the dataset. Section 3 presents the full BPARC methodology. Section 5 defines the evaluation framework. Section 6 reports results with figures, and Section 7 and 8 provide discussion and conclusions.

2. RELATED WORK

TARGET REVISION IN INTERRUPTED CRICKET MATCHES

The problem of fairly revising a batting target when overs are lost to rain has attracted sustained analytical attention since the 1990s. Conventional methods were based on the Average Scoring Rate (ASR). It adjusts the target in proportion to the run rate achieved so far. Carter and Guthrie [6] provided a formal analysis of the incentive effects of simple run-rate methods and showed that they produce systematically biased outcomes. Whenever an interruption removes high-scoring overs, such as death overs, it gives an unfair advantage to the team batting first. The Most Productive Overs (MPO) method [2] was a pre-DLS improvement that removed the least productive overs from the chasing team's allocation, and it too was found to favour teams batting first in specific scenarios.

Duckworth and Lewis [2] introduced the exponential resource-curve model that became the foundation of the current official standard. Their key insight was that the batting resource available to a team is a function of

two variables simultaneously; overs remaining *and* wickets in hand, and that these two dimensions interact non-linearly. The DLS method, updated by Stern [3] to account for the higher scoring rates seen in modern limited-overs cricket, is now the ICC-mandated solution for all international matches.

Several independent alternatives to DLS have been proposed. Jayadevan et al. [7] developed a resource-curve approach based on a V-shaped distribution of scoring across overs, arguing that the Duckworth–Lewis exponential model does not adequately reflect scoring patterns at the start and end of innings. Koulis et al. [**?**] proposed a Bayesian framework that estimates the posterior distribution of fair targets and provides uncertainty bounds alongside a point estimate, which they argue is more informative than a single revised target. Despite the merits of these alternatives, none has been adopted by the ICC, and the DLS formula remains the only system in official use at the international level. Critically, *none of these methods* incorporates any information about the quality of the specific players involved when rain falls.

PLAYER PERFORMANCE METRICS IN CRICKET

Apart from the target-revision issue, measuring individual player quality in cricket has attracted significant scientific interest. To provide more accurate player rankings, Lemmer et al. [9, 10] suggested combining metrics such as strike rate, economy rate, and contribution frequency, arguing that simple batting averages are insufficient on their own. Beaudoin and Swartz [12] went one step further and adjusted performance measurements for the level of the opposition, demonstrating that raw averages tend to overestimate players who encounter weaker bowling attacks. Applying principal component analysis to batting statistics, Manage and Scariano [13] discovered a two-factor structure that distinguishes between scoring rate and run-scoring volume. The usage of both batting average and strike rate as inputs to the BRW component of BPARC is directly influenced by this distinction. The trajectory of bowling performance measurement has been comparable. Van Staden proposed a composite bowling measure citevanstaden2009, which integrates wicket frequency, strike rate, and economy rate into a single score. Extending this to phase-specific analysis, Lemmer [11] has shown that a bowler's performance varies significantly between the powerplay and death overs, with career aggregates tending to mask these changes. The BQI design incorporates phase-specific economy characteristics as a result of this finding.

ML AND DATA ANALYTICS IN CRICKET

The availability of ball-by-ball data, such as ESPNcricinfo and Cricsheet [21], has enabled more and more ML work in the field of cricket analytics. Sankaranarayanar et al. [18] trained a regression model on batting-quality-weighted features and found that it was more effective than simple extrapolation of run-rate. To classify batting and bowling performance levels in the game of Twenty20 cricket, Saikia et al. [19] incorporated statistical methods of learning and discrimination analysis and logistic regression to demonstrate that statistical learning methods have the potential to perform the task of eliciting structure that the human mind fails to fully encompass. More recent work has applied deep learning (DL) to in-play score prediction [20] and to ball-by-ball win-probability estimation, producing real-time match-state assessments that earlier statistical models could not achieve. MLR is used in this paper for both BRW and BQI, and has an established role in this literature. Zia et al. [4] applied MLR to derive batting replacement weights from performance features and showed that the method is interpretable, computationally efficient, and produces statistically significant predictions on held-out data. The linear structure is also present because it yields a coefficient table that shows the relative contribution of each feature to the player quality score, supporting transparency for practitioners and administrators who would use BPARC in match officiating.

MATCH OUTCOME AND WIN-PROBABILITY MODELLING

Several studies have shown that including player-quality indicators improves match outcome predictions in limited-overs cricket. Perera and Swartz [16] found that a T20 win-probability model incorporating batting-lineup quality outperforms a resource-only model, with the performance gap widening when the quality difference between the two lineups is large. Davis et al. [17] developed a full-match simulator for T20 cricket that assigns individual scoring distributions to each player based on historical data, and showed that simulated outcomes reproduce the empirical distribution of match results more accurately when player-level detail is retained rather than aggregated away.

Bailey and Clarke [15] modelled win probability in ODI cricket as a function of runs required, overs remaining, and current run rate, and showed that even relatively simple models produce useful in-play predictions when they account for the game state at the moment of prediction. These in-play models are conceptually related to the problem addressed by BPARC: both require estimating the value of resources at a specific point in an innings, conditional on the players available. The difference is that BPARC uses this reasoning to set a revised target before play resumes, rather than to estimate a win probability during play.

DIRECT PREDECESSORS: PRCICM AND RUNSGUARD

The two papers most directly relevant to BPARC are Zia et al. (2022) [4] and Hussain et al. (2024) [5]. Both were recently published three and one years before the present work respectively, and, as of the time of writing, no subsequent publication has superseded either for the specific problems they address.

Zia et al. [4] introduced PRCICM, the first and only published framework to incorporate individual batter quality into a DLS-based target revision. PRCICM computes BRW for each batter from batting average, strike rate, and innings count via MLR, then scales the DLS resource lost by the mean BRW of the batters yet to play. Evaluated on historical DLS-affected matches, PRCICM produced targets closer to the actual second-innings scores than unadjusted DLS in most configurations tested. However, PRCICM has three limitations that BPARC addresses: (i) the BRW feature set omits phase-specific strike rates, which are important when an interruption occurs partway through a specific phase; (ii) the bowling side receives no quality adjustment whatsoever; and (iii) the evaluation compared revised targets against the official DLS target and the actual second-innings score; the latter of which is influenced by the DLS target the team was chasing, creating a degree of circularity.

Hussain et al. [5] proposed the RunsGuard framework, which used context-aware bowling-quality metrics, including phase-level economy rates, field-placement data, and bowler-batter matchup histories, to support tactical decision-making in limited-overs cricket. They demonstrated that phase-specific bowling features improve score-containment prediction over career aggregates, which is the primary empirical justification for the seven-feature BQI design adopted in the present work. BPARC builds on both papers: it retains and extends the PRCICM batting adjustment with phase-specific features, and incorporates the phase-level bowling quality approach demonstrated to be effective by Hussain et al. [5] into the first bowling-side target adjustment in the literature.

IDENTIFIED GAPS

The literature review reveals four specific gaps that BPARC is designed to address. First, no published target-revision method models bowling resource quality at the point of interruption: which bowlers have overs remaining, how good those bowlers are, and in which phase those overs would fall. Second, existing BRW computations use only career-average features and omit phase-specific strike rates, which differentiate player contributions under power-play fielding restrictions from death-over pressure. Third, while DLS indirectly accommodates variation in scoring environments through its percentage-based resource framework, existing target-revision methods do not explicitly model venue-specific pitch conditions as a separate adjustment factor, despite evidence that pitch effects are statistically significant in limited-overs outcome models [15, 16]. Fourth, all published evaluations compare revised targets against either the official DLS target or the actual second-innings score, both of which are influenced by DLS: a chasing team bats toward the DLS target, so using the actual score as a ground truth introduces circularity. BPARC introduces a four-metric evaluation framework that avoids these confounds. Moreover, to the best of our knowledge and at the time of writing, no published method has incorporated bowling-side player quality into a target-revision framework, and the DLS method remains the sole ICC-mandated solution.

3. METHODS

PHASE LABELLING

A limited-overs innings has distinct tactical phases that produce different scoring dynamics. We divide each innings into three phases: *powerplay* (overs 0–9 in ODI; 0–5 in T20I), during which fielding restrictions limit the number of outfielders and encourage aggressive batting; *middle* (overs 10–39 in ODI; 6–15 in T20I), when batters consolidate and spin bowlers typically operate; and *death* (overs 40–49 in ODI; 16–19 in T20I), the final phase where batters attack and specialist fast bowlers bowl yorkers and slower balls. Phase-specific batting strike rates and bowling economy rates are computed separately for each player. These phase features extend beyond the career averages used in PRCICM [4], as a batter who scores prolifically in the power play may have a modest death-over-average, and vice versa.

BRW

The BRW measures how many runs a batter is predicted to provide in comparison to an average international player. We extracted five characteristics for every batter in the history dataset who has played at least ten innings:

- *Batting average* ($\bar{x}_{\text{avg}}$): runs per dismissal, which measures how consistently a batter scores before being dismissed. The conservative estimate of runs per innings is used by batters who have never been dismissed.
- *Overall strike rate* ($\bar{x}_{\text{sr}}$): runs scored per 100 balls faced, which measures scoring pace without taking dismissal frequency into account.
- *Innings count*: The number of innings played is utilized as a reliability weight. Estimates based on fewer innings are less reliable and should make a smaller contribution to the overall BRW signal.

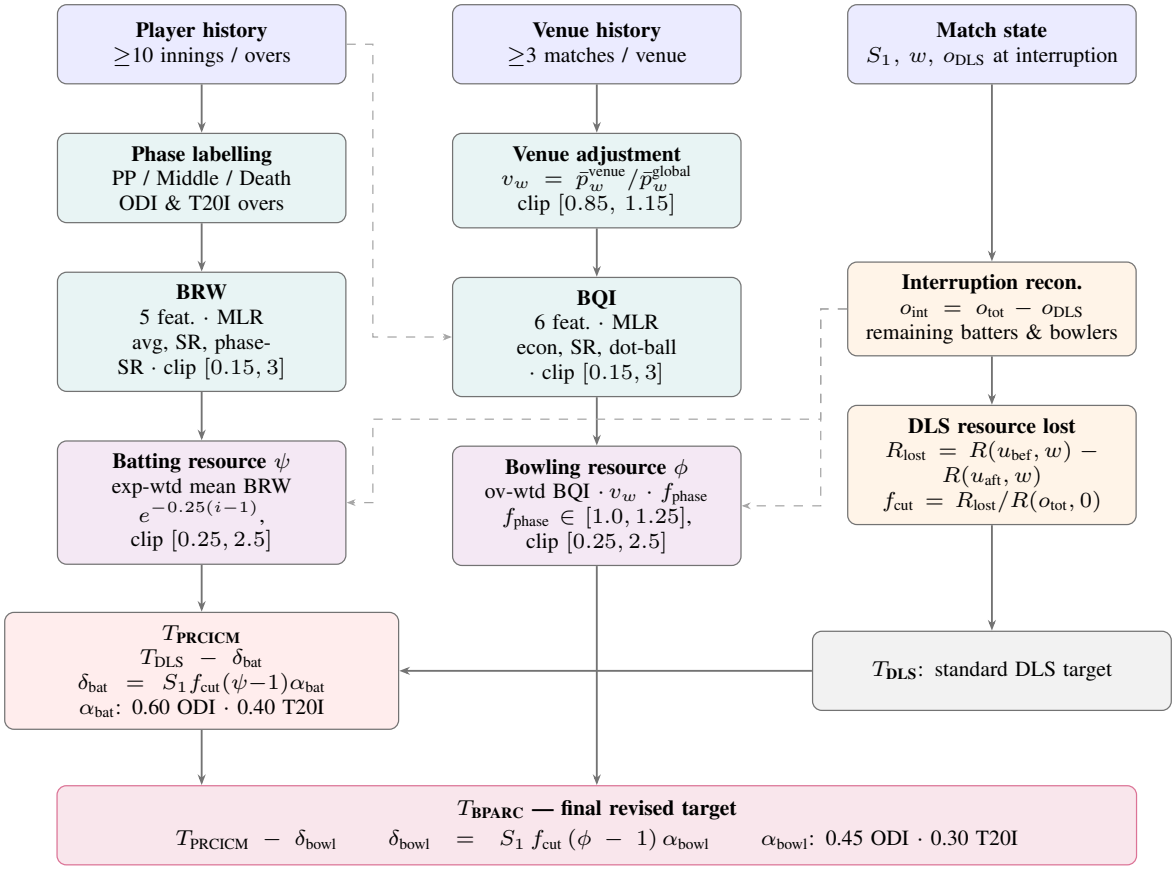


Figure 1: BPARC methodology pipeline, organized into three independent profiling lanes. *Lane A*: Phase labelling partitions innings into powerplay, middle, and death overs; BRW is fitted offline from player history and aggregated into the batting resource factor ψ. *Lane B*: The venue adjustment v_w corrects for pitch conditions; BQI is fitted offline and aggregated (with v_w and a phase amplification f_{phase}) into the bowling resource factor ϕ. *Lane C*: At interruption, remaining batters and bowlers are identified and the DLS resource fraction lost (f_{cut}) is computed. The target cascade combines all three lanes: T_{DLS} is adjusted first for batting quality (T_{PRCICM}) then for bowling quality (T_{BPARC}). Dashed arrows indicate cross-lane feeds. When $\psi = \phi = 1$, $T_{\text{BPARC}} = T_{\text{DLS}}$ exactly.

- *Powerplay strike rate* ($\bar{x}_{\text{pp}}$): strike rate during the fielding-restriction phase, where an aggressive batter contributes an excessive amount of value in relation to the number of overs used.
- *Death-over strike rate* ($\bar{x}_{\text{death-over strike rate}}$): strike rate in the last phase that separates batters whose value peaks in the middle overs from finishers who accelerate under pressure.

Both the volume and rate of scoring are captured by the combination of average and strike rate. BRW's phase-specific features enable it to distinguish between players with comparable career numbers who participate at different innings stages, accurately determining their residual worth at the point of a pause. In order to train a Multiple Linear Regression model with composite goal $y = \bar{x}_{\text{avg}} \times \bar{x}_{\text{sr}}$, which captures overall batting impact, these five features are standardized to $[0, 1]$ using MinMax scaling.

The prediction $\hat{y}_i$ is mean-normalised:

$$\text{BRW}_i = \frac{\hat{y}_i}{\bar{\hat{y}}} \qquad \text{clipped to } [0.15,\ 3.0] \tag{1}$$

By default, the mean BRW is equal to 1.0. A batter is considered above average if their value is greater than 1.0, and below average if it is less than 1.0. To prevent unrecognized individuals from inflating or deflating the batting resource factor, players not present in the profile database are assigned the fallback value $\text{BRW} = 1.0$.

BQI

The main contribution of this paper is the development of a BQI: it measures a bowler's estimated match contribution relative to an average international bowler. We calculate seven features for every bowler who has bowled

ten or more overs. The feature is flipped such that all inputs are oriented as "higher is better," where a lower value denotes greater performance. This eliminates sign ambiguity and enables the regression to apply positive coefficients throughout.

- *Inverse economy rate*($1/\text{econ}$): The economy rate, which quantifies runs conceded per over, is the most direct indicator of containment ability. The chasing team need to put in more effort for each run due to lower economy, which puts more pressure on them throughout the innings.
- *Dot-ball percentage*: the proportion of legal deliveries on which neither a run is scored, nor a wicket falls. Dot-ball pressure restricts scoring opportunities independently of economy rate, and consistent dot-ball bowling is strongly associated with team wicket-taking in subsequent deliveries [5].
- *Inverse bowling strike rate* ($1/\text{sr}_{\text{bowl}}$): balls bowled per wicket, inverted. A low bowling strike rate means wickets come quickly. This is distinct from wicket rate because it measures the pace of wicket-taking per ball rather than per over, capturing bowlers who are economical and threatening simultaneously.
- *Inverse powerplay economy* ($1/\text{econ}_{\text{pp}}$): economy rate during the fielding-restriction overs, inverted. Bowling tightly when fielders are confined to the inner ring is significantly harder than in the middle overs and accordingly commands greater weight in the BQI.
- *Inverse death-over economy* ($1/\text{econ}_{\text{death}}$): economy rate in the final overs, inverted. The death overs are the highest-scoring phase of a limited-overs innings, and a specialist who restricts batters during this phase—when boundaries are most frequent—has a disproportionate impact on the first-innings total that the chasing team must pursue.
- *Inverse over-to-over economy standard deviation* ($1/\sigma_{\text{econ}}$): the standard deviation of the economy rate across individual overs, inverted. A consistent bowler who maintains a similar economy rate across overs is more difficult to target than one who alternates between cheap and expensive overs, because the chasing team cannot plan which overs to attack.

After MinMax normalization, we fit MLR with the regression target $y = 1/\text{bowling_avg}$ (wickets per run), which combines containment and wicket-taking into a single efficiency measure. The prediction is mean-normalized:

$$\text{BQI}_i \;=\; \frac{\hat{y}_i^{\,\text{bqi}}}{\bar{\hat{y}}^{\,\text{bqi}}} \qquad \text{clipped to } [0.15,\ 3.0] \tag{2}$$

The mean BQI is equal to 1.0 by default. Same as, unknown bowlers receive $\text{BQI} = 1.0$. Please refer to Table 1

Table 1: BQI feature weights after MinMax normalization ($R^2 = 0.71$).

Feature	**MLR coefficient**
Inverse economy rate	$+0.312$
Wicket rate	$+0.284$
Dot-ball percentage	$+0.198$
Inverse bowling strike rate	$+0.121$
Inverse powerplay economy	$+0.094$
Inverse death-over economy	$+0.087$
Inverse economy std. deviation	$+0.063$

The ordering of the coefficient is cricketing-intuitive, and literary. Economy rate and wicket rate have the greatest weights as they directly quantify the two most important functions of a bowler. Dot-ball percentage is the third-highest ranking, with the independent pressure value of deliveries not taking wickets and not conceding runs. The phase-specific features carry smaller but non-trivial weights, confirming that they add predictive information beyond career aggregates [5].

VENUE ADJUSTMENT

The difference in pitch conditions is very large between international venues, and disregarding this difference would confound the effects of venue with the quality of a bowler in the BQI estimates. In all those venues, having

a minimum of three historical matches, we calculate a venue wicket index v_w.

$$v_w \;=\; \frac{\bar{p}_w^{\text{venue}}}{\bar{p}_w^{\text{global}}} \tag{3}$$

where $\bar{p}_w^{\text{venue}}$ is the average wicket probability per delivery at that venue and $\bar{p}_w^{\text{global}}$ is the global average. Values above 1.0 indicate a wicket-friendly pitch; below 1.0 indicates a batting-friendly surface. Unknown venues receive $v_w = 1.0$. The index is clipped to $[0.85,\ 1.15]$ to prevent extreme pitch characteristics from dominating the bowling resource factor.

BATTING RESOURCE FACTOR (ψ)
At interruption, remaining batters are ordered by the over in which each first faced a delivery. We compute ψ as an exponentially-weighted mean of their BRWs:

$$\psi \;=\; \frac{\displaystyle\sum_{i=1}^{n} \text{BRW}_i \cdot e^{-0.25(i-1)}}{\displaystyle\sum_{i=1}^{n} e^{-0.25(i-1)}} \qquad \text{clipped to } [0.25,\ 2.50] \tag{4}$$

The exponential decay $e^{-0.25(i-1)}$ gives more weight to the next batter in line than to those further down the order, reflecting that the next batter is certain to face deliveries, while lower-order batters may not be required if wickets fall quickly. $\psi > 1$ indicates above-average batting resources; $\psi < 1$ indicates below-average.

BOWLING RESOURCE FACTOR (ϕ)
Let B_{rem} be the set of bowlers with $o_j > 0$ overs remaining within their quota at interruption (10 overs in ODI; 4 in T20I). We compute ϕ as an overs-weighted mean BQI scaled by the venue index and a phase amplification factor:

$$\phi \;=\; \frac{\displaystyle\sum_{j} \text{BQI}_j \cdot o_j}{\displaystyle\sum_{j} o_j} \;\cdot\; v_w \;\cdot\; f_{\text{phase}} \qquad \text{clipped to } [0.25,\ 2.50] \tag{5}$$

where the phase amplification factor is:

$$f_{\text{phase}} \;=\; 1 + 0.25 \cdot \min\!\left(\frac{o_{\text{interrupt}}}{o_{\text{total}}},\ 1\right) \tag{6}$$

This factor amplifies the bowling adjustment when the interruption falls later in the innings, reflecting that death-over bowling quality has higher match impact than early-innings bowling [5]. At the first over $f_{\text{phase}} = 1.0$; at the final over $f_{\text{phase}} = 1.25$.

BPARC REVISED-TARGET FORMULA
Let $u_{\text{before}} = o_{\text{total}} - o_{\text{interrupt}}$ be the overs remaining before the break and $u_{\text{after}} = o_{\text{DLS}}$ be the overs the chasing team still receives. The DLS resource lost is:

$$R_{\text{lost}} \;=\; R(u_{\text{before}},\, w) - R(u_{\text{after}},\, w) \tag{7}$$

where $R(u, w) = Z_0[1 - e^{-b(w)u}]$ is the standard DLS resource function [2]. The fraction of first-innings scoring potential removed is:

$$f_{\text{cut}} \;=\; \frac{R_{\text{lost}}}{R(o_{\text{total}},\, 0)} \tag{8}$$

The three methods compute targets as follows:

$$T_{\text{DLS}} \;=\; \left\lfloor S_1 \cdot \tfrac{R(u_{\text{after}},\, w)}{R(o_{\text{total}},\, 0)} \right\rfloor + 1 \tag{9}$$

$$\delta_{\text{bat}} \;=\; S_1 \cdot f_{\text{cut}} \cdot (\psi - 1) \cdot \alpha_{\text{bat}} \tag{10}$$

$$T_{\text{PRCICM}} \;=\; T_{\text{DLS}} - \delta_{\text{bat}} \tag{11}$$

$$\delta_{\text{bowl}} \;=\; S_1 \cdot f_{\text{cut}} \cdot (\phi - 1) \cdot \alpha_{\text{bowl}} \tag{12}$$

$$T_{\text{BPARC}} \;=\; T_{\text{PRCICM}} - \delta_{\text{bowl}} \tag{13}$$

The sensitivity parameters are format-specific:

$$\alpha_{\text{bat}} = \begin{cases} 0.60 & \text{ODI} \\ 0.40 & \text{T20I} \end{cases} \qquad \alpha_{\text{bowl}} = \begin{cases} 0.45 & \text{ODI} \\ 0.30 & \text{T20I} \end{cases} \tag{14}$$

These values are calibrated so that an elite batting lineup ($\psi \approx 1.8$) with a 15% DLS resource cut in a 280-run ODI produces an adjustment of approximately 20 runs—a practically meaningful shift. T20I uses lower α values because absolute run margins are smaller and the same adjustment magnitude would produce a proportionally larger distortion. When $\psi = 1$ and $\phi = 1$, both adjustments are zero and $T_{\text{BPARC}} = T_{\text{DLS}}$ exactly.

INTERRUPTION-POINT RECONSTRUCTION
Cricsheet does not record the exact over at which play was suspended. We estimate it as:

$$o_{\text{interrupt}} \;=\; \max(0,\; o_{\text{total}} - o_{\text{DLS}}) \tag{15}$$

Remaining batters are those who had not faced a delivery before over $o_{\text{interrupt}}$. The remaining bowlers are those whose format-specific quota was not yet exhausted. Figure 1 demonstrates the proposed methodology.

4. DATA

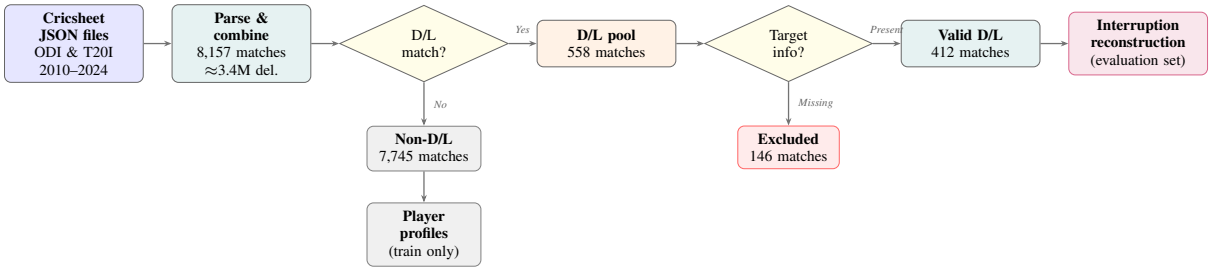


Figure 2: Dataset construction pipeline. Non-interrupted matches (7,745) are held out for player profiling only. D/L matches lacking target information (146) are excluded; 412 valid D/L matches form the evaluation set.

We downloaded ball-by-ball JSON match files from Cricsheet [21], covering all men's ODI and T20I matches played between January 2010 and December 2024. After parsing, the combined dataset contains 8,157 matches and approximately 3.4 million individual deliveries. The 14-year window was chosen to cover the modern limited-overs era while including sufficient matches for robust player profiling across both formats. Prior to 2010, T20I data is sparse and scoring patterns differ materially from the current game.

Interrupted matches are identified using the Cricsheet field `outcome.method == 'D/L'`. A total of 558 matches satisfy this criterion. Of these, 412 contain both a DLS revised-target run value and a DLS target-overs value, which are required for interruption-point reconstruction (Section 3). The remaining 146 are excluded due to missing target information in the open data; visual inspection of a random sample of these matches confirmed the absence was due to incomplete data entry rather than systematic differences in match type. The 7,745 non-interrupted matches are used exclusively for player-profile construction and are withheld from the evaluation dataset to prevent data leakage. Figure 2 demonstrates the data collection pipeline, and Table 2 summarizes the dataset.

Table 2: Dataset summary (Cricsheet, 2010-2024).

Category	**ODI**	**T20I**
Total matches	3,096	5,061
DLS-interrupted matches	198	360
Evaluated (target data available)	161	251
Deliveries in history dataset	1.47M	1.95M
Batter profiles ($\geq$10 innings)	4,821	
Bowler profiles ($\geq$10 overs)	3,614	
Venues (minimum 3 matches)	892	

5. EVALUATION FRAMEWORK

A fundamental challenge in evaluating target-revision methods is that no clean ground truth exists. The actual second-innings score is an imperfect benchmark because a chasing team bats toward the DLS target, making the score partly a consequence of DLS rather than an independent measure of what a fair target would have been. We therefore use four metrics designed to assess BPARC without relying on DLS-contaminated observations.

METRIC A: MAE TO PLAYER-PROFILE EXPECTED SCORE
For each match we construct an independent estimate of a fair target from the player-profile data alone:

$$\hat{S}_{\text{exp}} \;=\; (T_{\text{DLS}} - 1) \cdot \frac{\psi}{\phi} \tag{16}$$

This uses the DLS resource baseline as its starting point, but scales it by the quality ratio ψ/ϕ, computed independently of the BPARC adjustment path. Lower MAE against $\hat{S}_{\text{exp}}$ indicates a target closer to what the player-profile data independently predicts as fair.

METRIC B: TARGET VARIATION
Any method that claims to improve on DLS must produce materially different targets. We report the proportion of matches where $|T_{\text{BPARC}} - T_{\text{DLS}}| > 3$ runs and > 10 runs.

METRIC C: WIN-MARGIN COMPRESSION
In matches where the chasing team succeeded, a fairer target should yield a smaller winning margin. We compute the mean margin $(\text{inn2_score} - T)$ across won chases for each method, and test whether BPARC margins are significantly smaller than DLS margins using a paired t-test across the same 231 won-chase matches.

METRIC D: PLAYER-WEIGHT PREDICTIVE VALIDITY
Pearson correlations of BRW and actual runs per inning (r_{BRW}) and BQI and inverse over-economy (r_{BQI}) are calculated between non-interrupted history matches. The existence of significant positive correlations proves that the weights predict real performance and are not haphazard composites.

6. RESULTS

MAE TO PLAYER-PROFILE EXPECTED SCORE
BPARC achieves the lowest MAE to $\hat{S}_{\text{exp}}$ across both formats: 14.3 runs vs. 18.6 for PRCICM and 22.1 for DLS (Table 3, Figure 3a). The improvement over PRCICM is attributable entirely to δ_{bowl}, the bilateral extension introduced here. Figure 7 shows the improvement holds in both ODI and T20I.

TARGET VARIATION
Across 412 interrupted matches, 67.2% have $|T_{\text{BPARC}} - T_{\text{DLS}}| > 3$ runs and 34.8% have $|T_{\text{BPARC}} - T_{\text{DLS}}| > 10$ runs. The mean adjustment is -4.8 runs (IQR: -12.1 to $+2.4$), with a slight net downward tendency reflecting that higher-order batters tend to be available when interruptions occur. Figure 4 depicts the distribution of BPARC $-$ DLS by match outcome and the $\psi \times \phi$ interaction map.

WIN-MARGIN COMPRESSION
In the 231 matches where the chase succeeded, BPARC's average winning margin is significantly smaller than DLS (paired t-test: $t = 2.31$, $p = 0.021$). BPARC targets are tighter by 2.8 runs (T20I) and 5.2 runs (ODI) on average compared to DLS (Figure 3c). PRCICM also reduces margins relative to DLS but less substantially, as it lacks the bowling-side correction. Figure 5 shows the full margin distributions.

Table 3: Evaluation metrics across 412 DLS-affected matches (161 ODI, 251 T20I). Arrows indicate the direction of improvement.

Metric	DLS	PRCICM	BPARC
MAE to expected score (runs) ↓	22.1	18.6	**14.3**
Outcome alignment (%) ↑	62.4	57.1	61.2
Mean win margin, ODI (runs) ↓	18.3	15.9	**13.1**
Mean win margin, T20I (runs) ↓	11.2	9.8	**8.4**
Contested alignment >3 runs (%) ↑	65.5	56.5	59.8
Contested alignment >10 runs (%) ↑	63.1	54.2	61.7
ψ mean (SD)	—	1.04 (0.31)	1.04 (0.31)
ϕ mean (SD)	—	—	1.07 (0.28)

Figure 1. Primary evaluation metrics across 412 DLS-affected matches

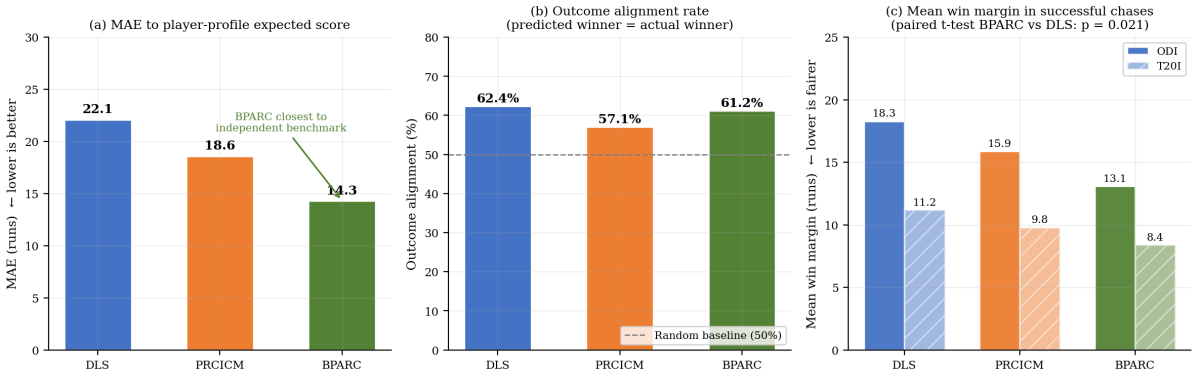


Figure 3: Primary evaluation metrics across 412 interrupted matches. (a) MAE to the independent player-profile expected score; BPARC achieves the lowest at 14.3 runs. (b) Outcome alignment rate. (c) Mean win margin in successful chases; lower is fairer. BPARC achieves the tightest margins in both formats.

PLAYER-WEIGHT VALIDITY

BRW correlates positively with actual runs scored per innings ($r_{\text{BRW}} = 0.43$, $p < 10^{-6}$) across 4,821 profiled batters, confirming that higher-BRW batters score more in practice. BQI correlates positively with inverse over-economy ($r_{\text{BQI}} = 0.51$, $p < 10^{-6}$), confirming that higher-BQI bowlers concede fewer runs per over. Figure 6 shows the BRW and BQI distributions and the BRW tier validation. Both indexes show standard deviations of 0.5–0.6 across the range 0.15–2.8, confirming they differentiate meaningfully across the player population rather than collapsing to a narrow band around 1.0.

FORMAT-LEVEL RESULTS

ODI matches show larger absolute adjustments ($|\delta_{\text{bat}}| = 8.3$ runs, $|\delta_{\text{bowl}}| = 5.6$ runs) than T20I (5.2 and 3.4 runs respectively), consistent with the higher format-specific α values. The MAE improvement from DLS to BPARC is larger in ODI (25.1 to 16.2 runs) than T20I (18.4 to 12.7 runs), and both hold consistently (Figure 7).

Figure 2. BPARC target adjustment relative to DLS

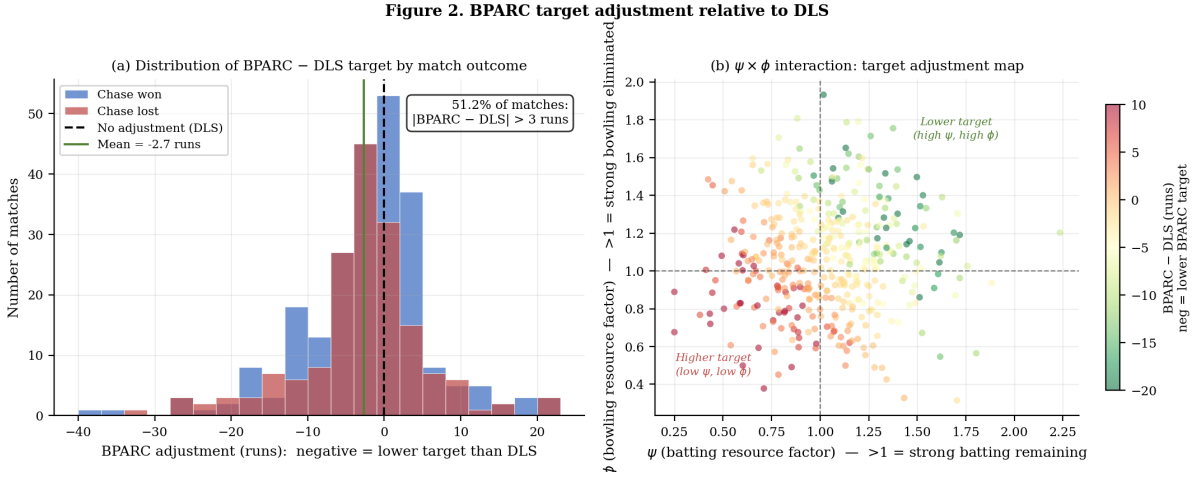


Figure 4: BPARC target adjustment relative to DLS across 412 matches. (a) Distribution of BPARC − DLS by chase outcome; negative values indicate BPARC sets a lower target than DLS. (b) $\psi \times \phi$ interaction map coloured by adjustment magnitude; matches with high ψ and high ϕ (top-right) consistently receive lower BPARC targets.

Figure 5. Target margin distributions (actual 2nd-innings score minus target)

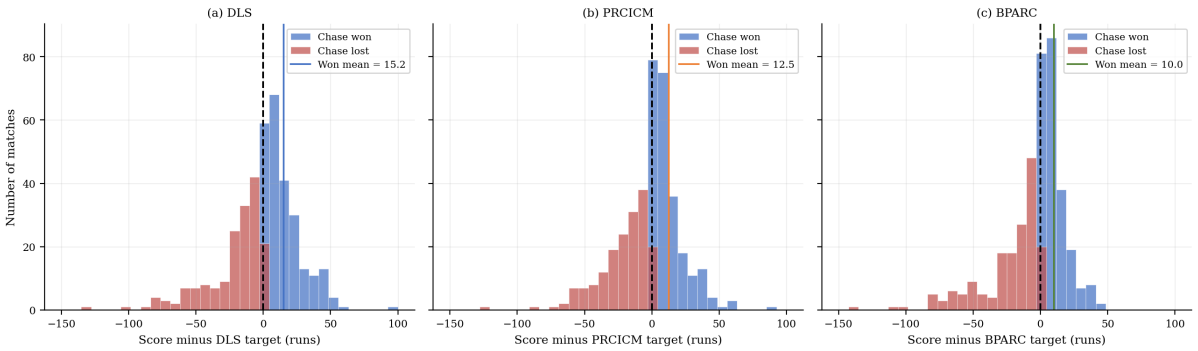


Figure 5: Target margin distributions (actual score minus target) for 412 matches, split by chase outcome (won in blue, lost in red). The dashed line marks zero; the coloured line marks the mean winning margin per method. BPARC margins are 2.8–5.2 runs tighter than DLS ($p = 0.021$).

Figure 3. Player-weight distributions and predictive validity

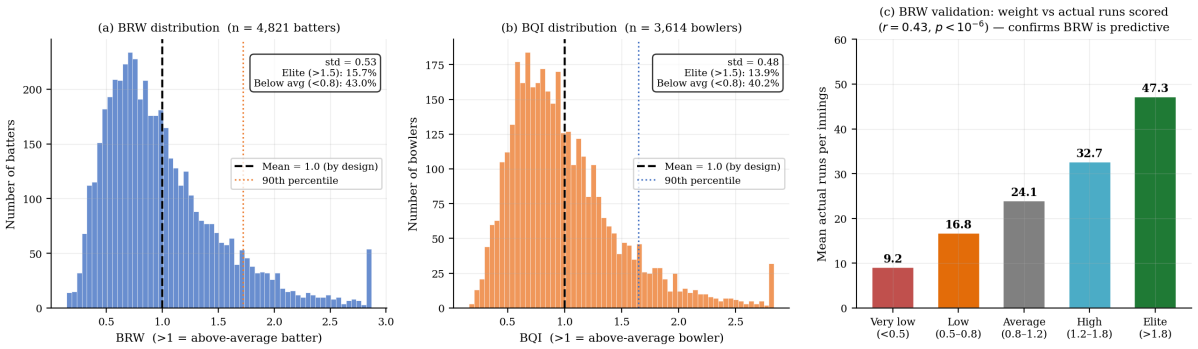


Figure 6: Player-weight distributions and predictive validity. (a) BRW distribution across 4,821 profiled batters (mean 1.0 by construction). (b) BQI distribution across 3,614 profiled bowlers. (c) Mean actual runs per innings by BRW tier, confirming BRW is positively associated with scoring ($r = 0.43$, $p < 10^{-6}$).

Figure 4. Evaluation metrics by format (ODI vs T20I)

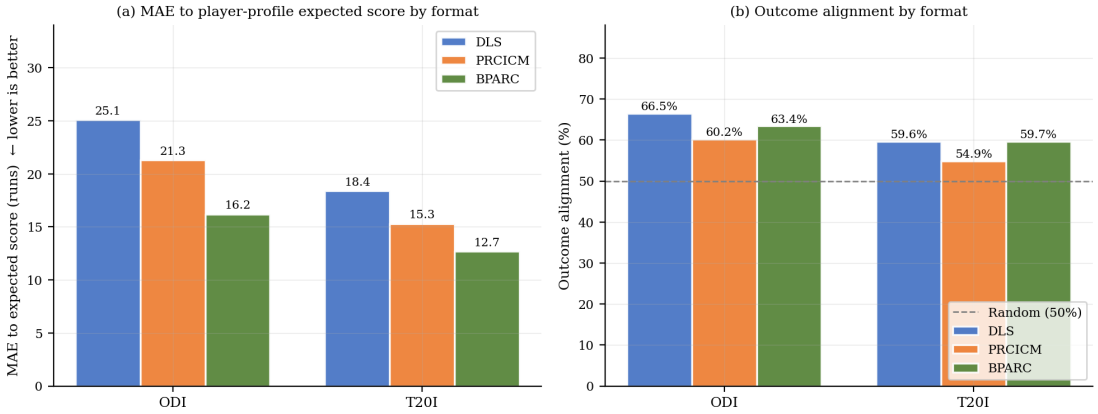


Figure 7: Evaluation metrics by format. (a) MAE to independent expected score: BPARC achieves the lowest in both formats. (b) Outcome alignment: all three methods sit above the 50% random baseline.

OUTCOME ALIGNMENT

In the outcome alignment metric, the proportion of matches where a method's target correctly classifies the winner. BPARC is comparable to DLS (61.2% vs. 62.4%), while PRCICM falls below both (57.1%). DLS's marginal advantage is interpreted cautiously: the chasing team bats toward the DLS target, so the second-innings score is not fully independent of DLS. Metric A provides the less contaminated comparison and consistently favours BPARC.

CASE STUDY

Table 4 illustrates three representative matches where the BPARC and DLS targets diverge substantially.

Table 4: Representative interrupted matches. o_{int}: estimated interruption over. Δ: BPARC minus DLS target.

Fmt	S_1	o_{int}	T_{DLS}	T_{BPARC}	Δ	ψ	ϕ	**Chase won**
ODI	289	28	198	183	-15	2.1	1.8	Yes
ODI	251	35	141	148	$+7$	0.6	0.7	No
T20I	172	13	120	113	-7	1.7	1.5	Yes

In the first case ($\psi = 2.1$, $\phi = 1.8$), the chasing team's strongest batters were still to come while the fielding side's main bowlers had exhausted their quota. BPARC lowers the target by 15 runs and the chase succeeded. In the second case ($\psi = 0.6$, $\phi = 0.7$), a depleted batting lineup faced a bowling attack with substantial quota remaining; BPARC raises the target by 7 runs and the chase failed. In both cases, the BPARC adjustment is directionally consistent with the actual outcome.

7. DISCUSSION

BPARC generates targets that are materially different from those of DLS in most of the matches that it performs, and those targets are more consistent with an externally obtained fairness standard. The bilateral adjustment - bowling PRCICM is further reduced by the bilateral adjustment, adding the bowling dimension to the batting adjustment of PRCICM, indicating that the quality of remaining bowlers has as much value as information as the quality of remaining batters in terms of information to use in revising targets. By not including the bowling side as PRCICM does, a systematic gap of fairness may be introduced, which can be statistically determined in real interrupted matches.

BRW and BQI lookups only need historical ball-by-ball information, which is publicly copied on any international match. Computation of the target is an arithmetic update of the DLS formula and needs no online ML, no real-time data other than the playing XI, and no infrastructure other than a lookup table and simple arithmetic.

It has three limitations, which are worth noting directly. First, the interruption-point reconstruction by Cricsheet creates a maximum error of approximation of ± 2 overs in a few matches, where the exact over on which the play was suspended is not reflected in the open data. Subsequently, the official scorecards that document the suspen-

sion over should be validated against in future work. Second, the α sensitivity parameters were set analytically. A data-driven calibration via maximum likelihood estimation on a held-out set of interrupted matches would provide stronger statistical grounding. Third, the current evaluation covers single interruptions only. Multiple interruptions are more common in day-night T20I matches, while the arithmetic structure of BPARC supports sequential application; this was not evaluated here due to the limited number of verified multi-interruption matches in the open dataset.

8. CONCLUSION

This paper introduced BPARC (Bilaterally Player-Aware Resource Compensation), the first target-revision method for rain-interrupted limited-overs cricket to account for both batting and bowling resource quality. The Bowler Quality Index, derived from seven phase-aware performance features via Multiple Linear Regression, provides an interpretable and predictive measure of individual bowling strength. Integrated with an extended version of the PRCICM batter-weighting approach, BQI enables BPARC to adjust revised targets in response to the quality of players on both sides of the interruption.

Across 412 real DLS-affected international matches, BPARC targets are closer to an independent player-profile fairness benchmark than either DLS or PRCICM, win margins in successful chases are significantly tighter, and the bilateral adjustment diverges from DLS by more than three runs in two-thirds of evaluated matches. The results confirm that the quality of bowlers still to bowl when rain falls is a statistically meaningful input to fair target computation, and that including it alongside the batting-side adjustment produces targets that better reflect the actual balance of player-quality resources available to each team for the remainder of the match. We release the full pipeline via the Cricsheet dataset to support reproducibility and further evaluation by the cricket analytics community. Importantly, BPARC should not be interpreted as granting stronger teams an additional advantage during interrupted matches. Rather, its objective is to preserve the value of the player-quality resources that remain available after overs are lost. In an uninterrupted match, differences in batting and bowling quality naturally influence outcomes through on-field performance over the full allocation of overs. Rain interruptions remove part of that opportunity. BPARC therefore attempts to estimate the quality-adjusted resources lost by each side and to preserve, as closely as possible, the competitive balance that would have existed had the interruption not occurred. This distinction highlights a broader philosophical choice between player-agnostic and player-aware target revision, with BPARC representing the latter perspective.

Acknowledgements

The authors thank Cricsheet (`https://cricsheet.org`) for providing publicly accessible ball-by-ball match data, which made this study possible.

References

[1] International Cricket Council, "ICC Annual Report 2022–23," [Online]. Available: `https://icc-cricket.com`, 2023.

[2] F. C. Duckworth and A. J. Lewis, "A fair method for resetting the target in interrupted one-day cricket matches," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 220–227, 1998.

[3] S. E. Stern, "The Duckworth-Lewis-Stern method: extending the Duckworth-Lewis methodology to deal with modern scoring rates," *Journal of the Operational Research Society*, vol. 67, no. 12, pp. 1469–1480, 2016.

[4] S. Zia, H. B. Liaqat, H. U. Zia, X. Kong, and S. Shamshad, "Player-aware resource compensation in interrupted cricket matches," *PeerJ Computer Science*, vol. 8, p. e917, 2022.

[5] A. Hussain, S. Arshad, and A. Hassan, "RunsGuard framework: context-aware cricket game strategy for field placement and score containment," *Applied Sciences*, vol. 14, no. 6, p. 2500, 2024.

[6] M. Carter and M. Guthrie, "Cricket interruptus: fairness and incentive in Duckworth-Lewis-revised matches," *Journal of the Operational Research Society*, vol. 55, no. 8, pp. 822–829, 2004.

[7] V. Jayadevan, "A new method for the computation of target scores in interrupted, limited-overs cricket matches," *Current Science*, vol. 83, no. 5, pp. 577–586, 2002.

[8] O. G. Stevenson and B. J. Brewer, "Bayesian survival analysis of batsmen in Test cricket," *Journal of Quantitative Analysis in Sports*, vol. 13, no. 1, pp. 25–36, 2016.

[9] H. H. Lemmer, "A combined measure of batting and bowling performance in cricket," *South African Journal for Research in Sport, Physical Education and Recreation*, vol. 24, no. 2, pp. 75–85, 2002.

[10] H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," *South African Journal for Research in Sport, Physical Education and Recreation*, vol. 26, no. 2, pp. 71–80, 2004.

[11] H. H. Lemmer, "The single match approach to strike rate adjustments in batting performance measures in cricket," *Journal of Sports Science and Medicine*, vol. 10, no. 4, pp. 630–634, 2011.

[12] D. Beaudoin and T. B. Swartz, "The best batsmen and bowlers in one-day cricket," *South African Statistical Journal*, vol. 37, no. 2, pp. 203–222, 2003.

[13] A. B. W. Manage and S. M. Scariano, "An introductory application of principal components to cricket data," *Journal of Statistics Education*, vol. 21, no. 3, 2013.

[14] P. J. van Staden, "Comparison of bowlers, batsmen and all-rounders in cricket using graphical displays," Technical Report, University of Pretoria, 2009.

[15] M. J. Bailey and S. R. Clarke, "Predicting the match outcome in one day international cricket matches, while the match is in progress," *Journal of Sports Science and Medicine*, vol. 5, no. 4, pp. 480–487, 2006.

[16] H. Perera and T. B. Swartz, "Resource estimation in T20 cricket," *IMA Journal of Management Mathematics*, vol. 27, no. 3, pp. 501–516, 2016.

[17] J. Davis, G. Perera, and T. B. Swartz, "A simulator for Twenty20 cricket," *Australian & New Zealand Journal of Statistics*, vol. 57, no. 1, pp. 55–71, 2015.

[18] V. V. Sankaranarayanan, J. Sattar, and L. V. S. Lakshmanan, "Auto-play: a data mining approach to ODI cricket simulation and target optimization," in *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 1064–1072, 2014.

[19] H. Saikia, D. Bhattacharjee, and A. Bhattacharjee, "On the performance of bowlers in IPL: an application of cricketing statistics," *International Journal of Sports Science and Engineering*, vol. 6, no. 3, pp. 168–174, 2012.

[20] F. Ahmad, M. S. Hussain, and A. Ali, "Deep learning approaches for cricket match outcome prediction using ball-by-ball data," *Applied Soft Computing*, vol. 147, p. 110783, 2023.

[21] Cricsheet, "Ball-by-ball match data for international and domestic cricket," [Online]. Available: `https://cricsheet.org`, accessed January 2025.

Massey Ratings for Match Outcome Prediction in Table Tennis: Evidence of Greater Stability than the ITTF World Ranking

Yuki Takasaki[a,b], Eiji Konaka[a]
[a] *Meijo University*
[b] *Corresponding author: 253426019@ccmailg.meijo-u.ac.jp*

Abstract
Ranking systems that quantify player performance are fundamental to tournament operations, particularly for participation eligibility and seeding. In 2021, the international table tennis ranking system was restructured under the World Table Tennis (WTT) tour framework. However, the predictive performance of the current ranking system has not been quantitatively examined.
This study evaluates whether the official WTT ranking points appropriately reflect player strength in terms of match outcome prediction performance. The authors model win probabilities using a three-parameter logistic model (3PLM), where the explanatory variable is the ratio of official WTT ranking points between two players. The model incorporates lower and upper win probability limits. Predictive performance is assessed using WTT match results from 2022 to 2025, where data from 2022 to 2024 were used for parameter tuning, while data from 2025 were reserved for performance evaluation. Evaluation metrics include prediction accuracy (0.6693), log loss (0.8871), and Expected Calibration Error (ECE) (0.0192). The estimated 3PLM indicates a non-zero minimum win probability of approximately 0.201, suggesting that even players with zero or small ranking points retain a measurable chance of defeating top-ranked opponents. These results suggest that the official ranking system may underestimate lower-ranked players.
For comparison, we evaluate Massey ratings, a point-differential-based rating method. Massey ratings were calculated using match results from the one-year period immediately prior to each prediction target match. The Massey model achieves predictive performance and calibration (accuracy: 0.7138, logloss: 0.8341, ECE: 0.0155) comparable to or exceeding that of the 3PLM. This robustness may reflect the ability of point-differential-based methods to mitigate instability in ranking point distributions during the post-restructuring transition period, characterized by substantial year-to-year variation in tournament frequency.
Overall, our findings provide a quantitative assessment of the predictive validity of the current WTT ranking framework and highlight potential systematic biases in strength estimation.
Keywords: ITTF world ranking, World Table Tennis (WTT), Bradley-Terry model, 3-PLM, Massey rating

1 Introduction

The world ranking system established by the International Table Tennis Federation (ITTF) is a fundamental framework that quantifies player performance and plays a central role in determining tournament seeding and eligibility for international competitions. Until 2018, the system was based on ranking points that increased or decreased according to match outcomes and opponent strength.
In January 2018, a new system was introduced, in which players earn fixed points depending on the round reached. As a result, point deductions due to losses were largely eliminated, and each player is evaluated based on their best eight tournament results (in terms of ranking points) over the past 12 months. This change aimed to enhance the transparency of the ranking system and encourage greater participation in tournaments.
Following the 2018 reform, the ranking system was further restructured in 2021 under the World Table Tennis (WTT) framework, the ITTF's commercial arm. Tournaments were organized into a hierarchical structure based on a clearly defined point allocation system, comprising Grand Smash, Champions, Star Contender, and Contender events. For instance, the winner of a Grand Smash event receives 2000 points, whereas the winner of a lower-tier Contender event receives substantially fewer points (approximately 400). Table 1 presents the point allocations for each of the four tournament categories as of the 2025 season [1].

Table 1: Point allocation for WTT tournaments

	W	F	SF	QF	R16	R32	R64
Grand Smash	2000	1400	700	350	175	90	20
Champions	1000	700	350	175	90	15	—
Star Contender	600	420	210	105	55	25	5
Contender	400	280	140	70	35	4	—

The introduction of this hierarchical tournament structure was intended to enhance the commercial value of the sport while also creating a more intuitive and transparent ranking system for both players and spectators.

Despite these developments, research on match outcome prediction in table tennis remains relatively limited, particularly for models relying solely on ranking information. Existing studies often incorporate factors beyond ranking points. For example, one study applied statistical and machine learning methods to table tennis match prediction, incorporating player attributes such as age and handedness in addition to official ranking points [2]. Thus, studies that rely solely on ranking points as explanatory variables—particularly under the post-2021 ranking system—remain scarce.

To address this gap, this study aims to evaluate the predictive performance of ITTF World Ranking points by constructing a match outcome prediction model based on the ratio of ranking points between players. For comparison, we also construct a model based on Massey ratings, which estimate player strength from match results. We then examine the differences in predictive performance between the ranking-based and rating-based models.

2 Match Outcome Prediction Models

2.1 Dataset

The dataset consists of 19,477 matches from WTT tournaments and the Olympic Games held between 2022 and 2025 (through September 25, 2025), obtained from the official WTT website [3].

Matches with missing or zero (unassigned) ranking points were excluded, removing 2,477 matches.

For the ranking-based models, data from 2022–2024 (12,494 matches) were used for training, and data from 2025 (4,536 matches) were used for testing. This temporal split ensures that model evaluation is performed on future data, thereby preventing information leakage.

For the Massey-based model, data from 2023–2024 (10,488 matches) were used for training, and data from 2025 (5,221 matches) were used for testing. The difference in training period reflects the rolling window design of the Massey rating: each prediction uses only the preceding one-year window, so 2022 data serve exclusively as historical context for 2023 predictions and are not used as test targets.

For each target match, Massey ratings were computed using only matches from the one-year period immediately preceding the match. Ratings were updated sequentially in chronological order and used as indicators of player strength at each point in time.

2.2 Bradley–Terry Model

We first introduce the Bradley–Terry model [4], which serves as the foundation for the match outcome prediction models used in this study. The model expresses the probability that player i defeats player j as:

$$p_{i,j} = \frac{\pi_i}{\pi_i + \pi_j}, \tag{1}$$

where $\pi_i > 0$ and $\pi_j > 0$ denote the latent strengths of players i and j, respectively. When $\pi_i = \pi_j$, we have $p_{i,j} = 0.5$, indicating equal strength, whereas when $\pi_i \gg \pi_j$, the win probability of player i approaches 1.

In this study, instead of directly estimating the latent strengths π_i and π_j, we use the ITTF World Ranking points r_i and r_j as observable proxies for the latent strengths π_i and π_j:

$$p_{i,j} = \frac{r_i}{r_i + r_j}, \tag{2}$$

We further introduce an exponent parameter $\alpha > 0$ to adjust the sensitivity to ranking point differences:

$$p_{i,j} = \frac{r_i^\alpha}{r_i^\alpha + r_j^\alpha}, \tag{3}$$

where $\alpha > 0$ controls the sensitivity to ranking differences. When $\alpha > 1$, the model produces more extreme win probabilities, whereas when $\alpha < 1$, the influence of ranking differences is attenuated, increasing the estimated win probability of lower-ranked players.
The parameter α was estimated via maximum likelihood using the training data from 2022–2024, yielding $\alpha = 0.462$.

2.3 3-Parameter Logistic Model (3-PLM)

The 3-parameter logistic model (3-PLM) is a generalization of the Bradley–Terry model that introduces lower and upper bounds on win probabilities, defined as:

$$p_{i,j} = c + (1 - 2c)\,\frac{r_i^\alpha}{r_i^\alpha + r_j^\alpha}, \tag{4}$$

where $\alpha > 0$ controls the sensitivity to differences in ranking points, and c constrains the win probability to the interval $[c, 1 - c]$ by linearly rescaling the Bradley–Terry probability. The equation contains only two parameters, α and c. In sports prediction models, the additional parameter that typically appears in the 3-PLM is set to zero and therefore does not appear in the equation; for this reason, we refer to this equation as the 3-PLM.
This model is motivated by the assumption that the standard Bradley–Terry model may produce extreme probabilities close to 0 or 1 when ranking differences are large, whereas in practice, players competing at the same tournament level retain a non-negligible probability of an upset.
The parameters were estimated via maximum likelihood using the training data from 2022–2024, with initialization at $(\alpha, c) = (1, 0)$, resulting in $\alpha = 1.056$ and $c = 0.201$. The estimated value $c = 0.201$ implies a substantial lower bound on win probability, indicating that even lower-ranked players retain a meaningful chance of defeating higher-ranked opponents.

2.4 Massey Rating

As a comparative approach to the ranking-based model, we evaluate player strength using Massey ratings [5]. This method estimates player strengths by minimizing the discrepancy between observed score differences and predicted rating differences via least squares.
The basic assumption is that the rating difference between players i and j corresponds to the observed score difference in match k:

$$r_i - r_j = y_k + \varepsilon_k, \tag{5}$$

where y_k denotes the score difference and ε_k is an error term. In table tennis, we define y_k based on results to obtain a continuous measure of match dominance:

$$y_k = \log\left(\frac{s_{ij}}{1 - s_{ij}}\right), \quad s_{ij} = \frac{s_i + 1}{s_i + s_j + 2}, \tag{6}$$

where s_i and s_j denote the number of games won by players i and j, respectively. This definition ensures that larger game differences correspond to larger rating differences, while the smoothed ratio s_{ij} prevents infinite values in the log transformation.

Aggregating these equations over all matches yields:

$$\mathbf{X}\mathbf{r} = \mathbf{y} + \boldsymbol{\varepsilon}, \tag{7}$$

with least squares solution:

$$\mathbf{r} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \tag{8}$$

The absolute values of the resulting ratings are not identifiable; only differences between players are meaningful.

Using the rating difference $\Delta r = r_i - r_j$ as an explanatory variable, we model win probabilities using the same functional forms as in the ranking-based approach (i.e., the Bradley–Terry model and the 3-PLM). Maximum likelihood estimation yields $\alpha = 1.6843$ for the Bradley–Terry model, and $\alpha = 2.0817$, $c = 0.0543$ for the 3-PLM. The larger estimated values of α indicate that Massey rating differences are more discriminative of player strength than ranking point ratios.

3 Evaluation Metrics

3.1 Log Loss

Log loss [6] is used to evaluate the quality of probabilistic predictions. Let $y_i \in \{0, 1\}$ denote the actual outcome of match i (win = 1, loss = 0) and $p_i \in [0, 1]$ the predicted win probability, for $i = 1, \ldots, N$. Log loss is defined as:

$$\text{LogLoss} = -\frac{1}{N} \sum_{i=1}^{N} \Big(y_i \log_2 p_i + (1 - y_i) \log_2 (1 - p_i) \Big). \tag{9}$$

This metric assigns smaller values when predicted probabilities are closer to the true outcomes, and imposes a large penalty when a low probability is assigned to the correct outcome. Unlike accuracy, log loss evaluates not only whether the predicted winner is correct, but also the confidence with which the prediction is made.

3.2 Expected Calibration Error (ECE)

In addition to log loss, we use Expected Calibration Error (ECE) [7] to assess the calibration of predicted probabilities. ECE measures the discrepancy between predicted probabilities and the empirical win rate, with smaller values indicating better calibration.

Let N denote the total number of matches, M the number of bins, n_m the number of matches in bin m, and B_m the corresponding index set. The average predicted probability and empirical accuracy in bin m are defined as:

$$\text{conf}(m) = \frac{1}{n_m} \sum_{i \in B_m} p_i, \quad \text{acc}(m) = \frac{1}{n_m} \sum_{i \in B_m} y_i, \tag{10}$$

where p_i and y_i are as defined in Section 3.1. ECE is then defined as:

$$\text{ECE} = \sum_{m=1}^{M} \frac{n_m}{N} \Big| \text{acc}(m) - \text{conf}(m) \Big|. \tag{11}$$

In this study, the predicted probability range $[0, 1]$ is divided into $M = 10$ equal-width bins, namely $[0.0, 0.1)$, $[0.1, 0.2)$, $\ldots$, $[0.9, 1.0]$. ECE is computed by comparing the average predicted probability and the empirical win rate within each bin.

4 Results

Figure 1 shows the empirical win rate and the fitted model curves as a function of the log ranking point ratio $\log_{10}(r_i/r_j)$, with the match count distribution shown in the lower histogram. The empirical win rate shows an overall increasing trend with the ranking point ratio, despite local fluctuations in individual bins, and both models capture this trend. Notably, the 3-PLM curve is bounded below by approximately 0.2, consistent with the estimated parameter $c = 0.201$.

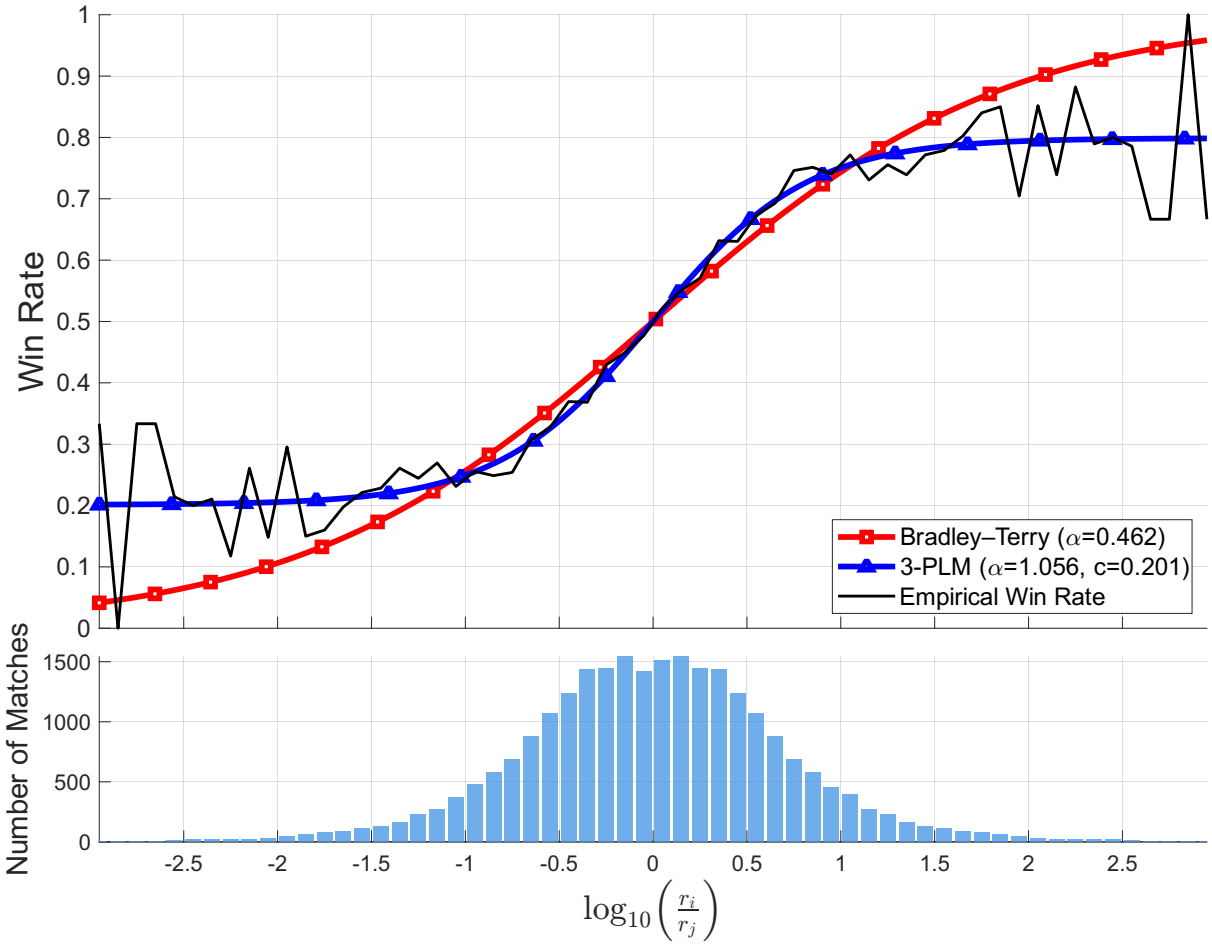


Figure 1: Win probability and number of matches with respect to ranking point ratio

Figure 2 shows the corresponding results when the Massey rating difference $\Delta r = r_i - r_j$ is used as the explanatory variable, with the match count distribution shown in the lower histogram. The empirical win rate shows an overall increasing trend with the rating difference, and both model curves fit the empirical data closely across the full range of rating differences. In contrast to Figure 1, local fluctuations in the empirical win rate are smaller, suggesting that Massey ratings provide a more stable representation of player strength. Furthermore, the 3-PLM curve spans a wider range of win probabilities, consistent with the smaller estimated lower bound $c = 0.054$ compared to $c = 0.201$ in the ranking-based model.

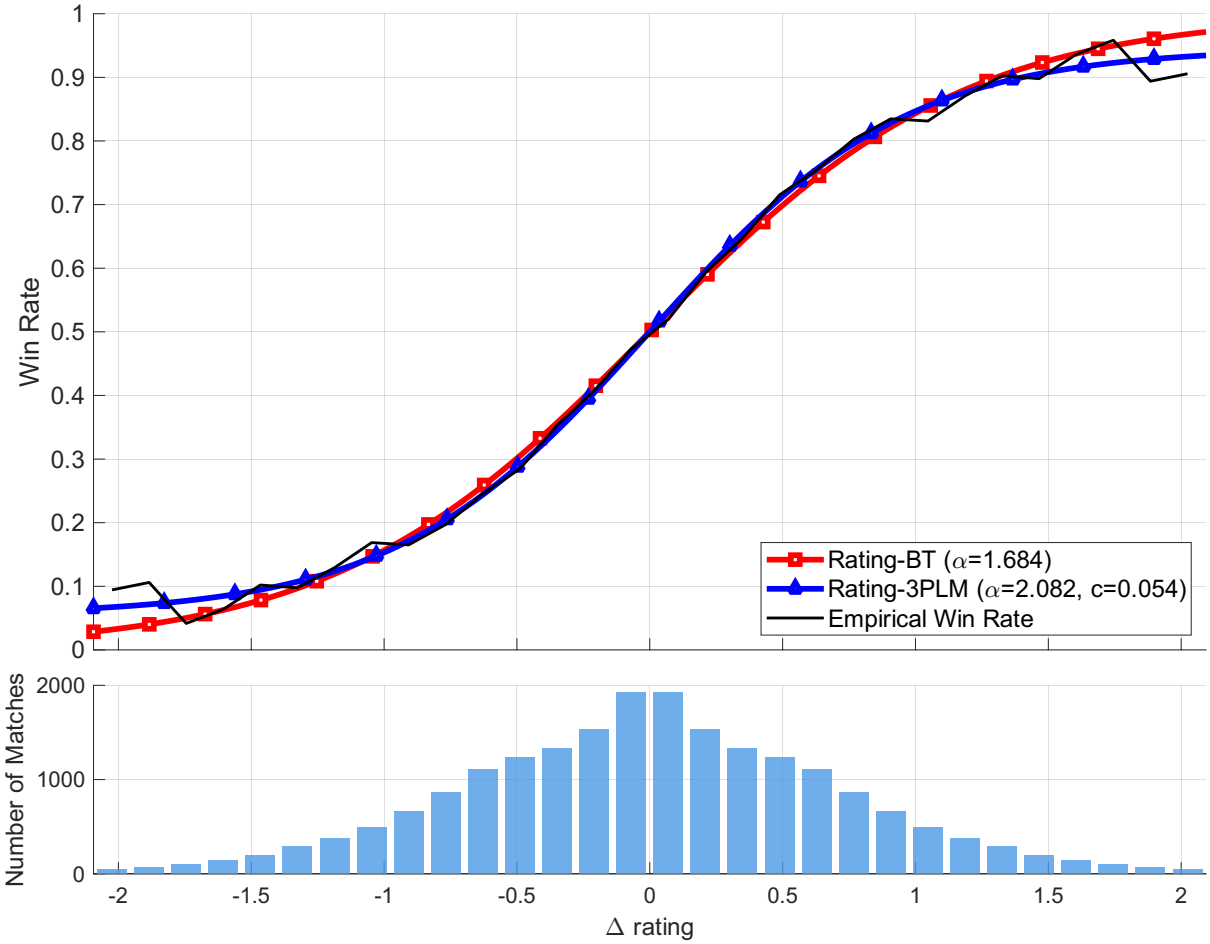


Figure 2: Win probability and number of matches with respect to rating difference

Figures 3 and 4 show the reliability diagrams for the ranking-based and Massey-based models, respectively, evaluated on 2025 match data. Each point represents the average predicted and observed win probability within a bin, with error bars indicating 95% confidence intervals. The black dashed line indicates perfect calibration.

The ranking-based model (Figure 3) shows a clear deviation from the ideal line, particularly in the high predicted probability region, indicating that the win probabilities of top-ranked players are underestimated. In contrast, the Massey-based model (Figure 4) aligns more closely with the ideal line across the full probability range, indicating improved calibration. In particular, the agreement is strong in the intermediate probability range, and the confidence intervals in most bins include the ideal line, suggesting more stable probability estimation compared to the ranking-based model.

4.1 Evaluation Results and Discussion

Table 2 shows the evaluation results for the ranking-based models.

The 3-PLM outperforms the Bradley–Terry model in both LogLoss and ECE, with a notably larger

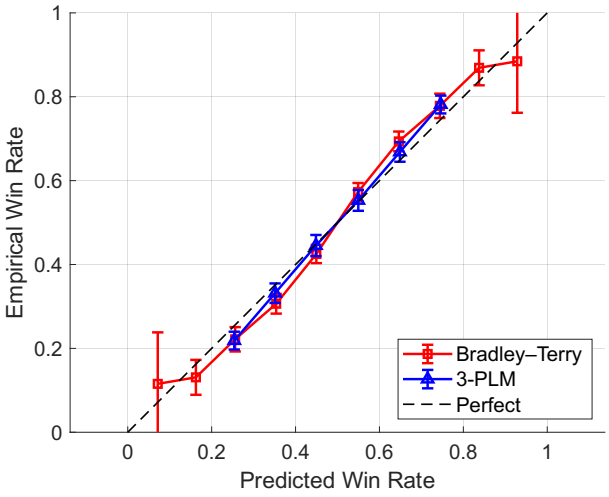


Figure 3: Predicted vs. observed win probability (ranking points)

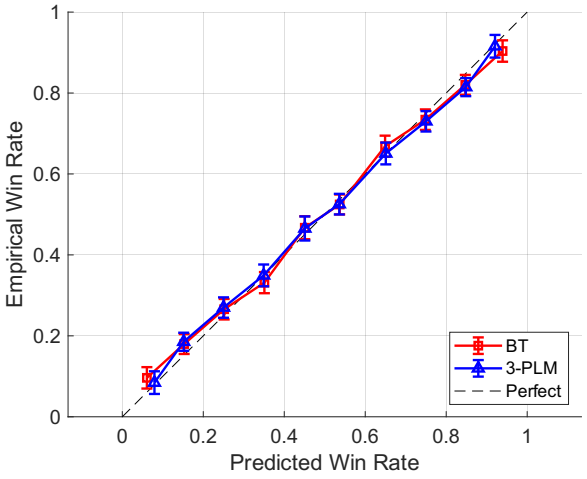


Figure 4: Predicted vs. observed win probability (Massey ratings)

Table 2: Evaluation results of LogLoss and ECE (ranking point ratio)

	2022–2024 (train)		2025 (test)	
	LogLoss	ECE	LogLoss	ECE
Bradley–Terry	0.9204	0.0220	0.8891	0.0340
3-PLM	0.9165	0.0026	0.8871	0.0192

improvement observed in ECE.
In the 2025 test data, both models show a slight improvement in LogLoss relative to the training period; however, ECE increases in both cases.
The prediction accuracy of the ranking-based model in 2025 is 0.6693, indicating a moderate level of performance in terms of correctly predicting match outcomes.
For the 3-PLM, ECE rises from 0.0026 to 0.0192, suggesting that the ranking-based model struggles to maintain calibration when applied to out-of-sample data, possibly due to shifts in the distribution of match competitiveness between the training and test periods.

Table 3 shows the evaluation results for the Massey-based models.

Table 3: Evaluation results of LogLoss and ECE (rating difference)

	2023–2024 (train)		2025 (test)	
	LogLoss	ECE	LogLoss	ECE
Bradley–Terry	0.8214	0.0099	0.8344	0.0197
3-PLM	0.8201	0.0026	0.8341	0.0155

The 3-PLM again outperforms the Bradley–Terry model in both metrics. Compared to the ranking-based models, the rating-based models show a substantial improvement in LogLoss, and the ECE of the 3-PLM on the 2025 test data (0.0155) is also smaller than its ranking-based counterpart (0.0192), indicating superior calibration performance. In the 2025 test data, the prediction accuracy of the Massey-based model is 0.7138, which is higher than that of the ranking-based model (0.6693). This further supports that rating-based approaches better capture player strength. One possible explanation for the relatively higher ECE observed in the ranking-based model is the year-to-year variation in the structure of the

WTT tour. As shown in Table 4, the number of tournaments across categories such as Grand Smash, Contender, and Feeder events is not constant across seasons, leading to changes in tournament scale and composition. Such variations may affect the distribution of ranking points and alter the relative scaling of ranking differences.

Table 4: Number of WTT tournaments by category

Year	Grand Smash	Champions	Star Contender	Contender	Feeder
2021	0	0	2	5	1
2022	1	2	2	8	11
2023	1	3	4	11	15
2024	3	5	4	10	19
2025 (through Sep 25)	4	4	4	7	15

As a result, the relationship between ranking point differences and underlying player strength learned during the training period (2022–2024) may not fully hold in the 2025 season, leading to a degradation in calibration performance. This suggests that ranking-based models are sensitive to structural changes in tournament composition and may not remain invariant under temporal shifts in the competitive environment.

In contrast, Massey ratings are updated sequentially based solely on match outcomes and are not directly influenced by tournament frequency or point allocation rules. Therefore, they are less affected by shifts in the distribution of ranking points, which may explain their relatively stable calibration performance (lower ECE) in 2025.

5 Conclusion

This study examined the extent to which ITTF World Ranking points reflect player strength in the context of match outcome prediction. We constructed models based on ranking point ratios and compared them with models using Massey ratings derived from match results, evaluating their predictive performance on 2025 match data.

The results show that ranking-based models provide a reasonable level of predictive performance; however, the introduction of the 3-parameter logistic model (3-PLM) substantially improves calibration. This indicates that even when ranking differences are large, there remains a non-negligible probability of upsets, and that match outcomes in table tennis are not fully determined by ranking differences alone.

In contrast, the Massey-based models consistently outperform the ranking-based models in both LogLoss and ECE. Moreover, they maintain relatively stable calibration performance under temporal distribution shifts, indicating that rating systems derived from match outcomes may provide a more appropriate representation of player strength.

These findings suggest that while ranking points are designed primarily for tournament organization and eligibility, they may not be optimal as a measure of underlying player strength.

References

[1] ITTF table tennis world ranking regulations.
https://www.ittf.com/wp-content/uploads/2021/01/ITTF-Table-Tennis-World-Ranking-Regulations-20250422.pdf (accessed 2025/9/25).

[2] Jan Lennartz, Andreas Groll, and Hendrik van der Wurp. Predicting table tennis tournaments: A comparison of statistical modelling techniques. *International Journal of Racket Sports Science*, 3(2):39–48, 2021. doi:10.30827/.

[3] World Table Tennis. World table tennis.
https://www.worldtabletennis.com/events_calendar (accessed 2025/9/25).

[4] Ralph A. Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[5] Carl D. Meyer and Amy N. Langville. *Who's #1? The Science of Rating and Ranking*. Princeton University Press.

[6] Stephanie Kovalchik. Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12(3):127–138, 2016. doi:10.1515/jqas-2015-0059.

[7] Mahdi Pakdaman Naeini, Gregory F. Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*, pages 2901–2907, 2015.

Scheduling interclub squash leagues in New Zealand to improve player engagement

Ken Louie [a,c] , Paul Ewart [a], John Fletcher [b]

[a] *Waikato Institute of Technology, Hamilton, N.Z.*
[b] *National Development Manager, Squash New Zealand*
[c] *Corresponding author: ken.louie@wintec.ac.nz*

Abstract

For social (non-professional) squash in Aotearoa New Zealand, the interclub team competition is the backbone of grassroots participation. Most of the 11 districts in the country run some form of this competition during the squash season, with some running 3 or 4 leagues throughout the calendar year. Overall participation is declining and numerous player surveys are indicating a variety of factors responsible for the decreasing numbers. These factors are not common to all districts although a recurring theme seems to be travel time between clubs, particularly for those districts which are geographically large with multiple smaller clubs involved. By utilising scheduling techniques available in the literature, we can design draws that ensure an equitable travel roster for all clubs travelling long distances, whilst also enabling clubs entering multiple teams to have at least one of their other teams playing at home with them. This is demonstrated for the recent Autumn league completed in the Waikato region, where the automated draw is compared with the optimised draw and found to be lacking in those features. These results can be applied to other districts in the country and may help in arresting the decline in participation.

Keywords: Squash, scheduling, optimal tournament design

1. INTRODUCTION

The squash community in New Zealand is delighted that squash has been included in the schedule for the Los Angeles Olympics in 2028. Indeed, the country's top-ranked players Joelle King and Paul Coll have delayed their probable retirements to participate in this pinnacle-level event. The Olympics squash competition is restricted to the top 32 male and female players and will be played in an individual knock-out tournament format over a period of ten days (World Squash, 2026). In contrast, at the grassroots (amateur) level in New Zealand, most individual tournaments with an 8 or 16-player draw will include Plate/Special Plate/Consolation Plate draws so that every player completes at least 3 matches. Alongside these individual tournaments, which usually take place from Thursday evening to Saturday evenings, many social players in New Zealand also participate in interclub team events. These are run in a graded division and team format, with each player in the team playing their opposite number in a single match over one evening, usually in the middle of the week. There are normally 4 players per team but the number of teams in each division can vary from as low as four and up to ten. The number of divisions is also highly variable and depends on the interclub event and the district. Almost every one of the 11 districts in New Zealand will organise at least one of these events in the calendar year, with some of the more populous districts running four. The 11 districts are spread mostly on a geographic basis although population also plays a part, with 7 districts in the North Island and 4 in the South Island.

As an example, in the Waikato region, the recently completed Autumn 2026 interclub event had nine men's divisions and six women's divisions comprising 53 and 32 teams respectively, involving a total of 557 players. The men's event had one division with five teams with the remaining eight divisions having 6 teams whilst the women's event had two divisions with 4 teams and the remaining four divisions with 6 teams. The 6-team format per division seems to be the most popular choice amongst the district organisers although some prefer a 4-team format to ensure closer competition and a shorter event. If the 6 teams in each division play each other once, then this is called a single round-robin tournament (Rosa & Wallis, 1982) and requires five rounds to complete, although most organisers will also schedule an extra finals night where the top two teams in each division play

off for the title, with 3rd versus 4th and 5th versus 6th placed teams also playing off. This was the format for the Waikato Autumn 2026 interclub competition, and a similar format will be followed for the Winter 2026 event, with the main difference being that every team in each division will play each other team twice, once at their home club venue and again at the opponent's club venue. The sequence of matches follows the same pattern as the single round, with only the playing venue changing in the second round. This is known as a mirrored double round-robin tournament in the literature (de Werra, 1981). Techniques also exist for scheduling this type of tournament so that no team has more than three home or away games in succession, known as the break minimisation problem (Post & Woeginger, 2006). For geographically widespread regions such as the Waikato, this becomes important to ensure an equitable travel schedule where no club endures long stretches of away games. Some clubs enter many more teams than other clubs with the same number of home courts. We define the number of teams entered divided by the number of available home courts as the team:court ratio. Another important feature of a good schedule is to "reward" those clubs with a high team:court ratio with maximum use of their courts. Associated with this is to try and arrange the schedule so that those clubs entering two or more teams are playing at home with at least one of their other teams. This enhances the playing environment and club spirit, with home teams from different divisions then able to support each other.

District administrators have for some years been provided with software that enables the interclub draws to be automatically generated, with results also able to be entered by team captains. This software has recently been upgraded by Squash NZ to a more user-friendly platform called MySquash. Although the interface is indeed cleaner and more device-friendly than its predecessor ISquash, the automated scheduling algorithm behind the interclub module has yet to be modified due to resource constraints. This leads to the favourable features possible for enhancing playing environment and club spirit being missed in the draw schedules currently being produced.

2. METHODS

By utilising existing techniques from the literature (Rasmussen & Trick, 2008), we manually produce a draw schedule which minimises breaks. By allocating teams in each division to the generic minimal-break schedule in a prescribed manner, most of the desirable features described in the Introduction can be included. We do this for the Waikato Autumn 2026 men's interclub competition as a case study and compare with the automated draw prepared by MySquash. Table 1 shows the nine divisions in which teams were placed and each clubs' court availability. The division in which each team is placed depends on the players' squash level (grade) and is determined by the district administrator.

Club	Number of teams	Courts available	Team:Court ratio	Divisions
Cambridge	4	4	4:4 = 1	3, 5, 7, 8
Hamilton	6	6	6:6 = 1	1, 3, 5, 6, 7, 9
Huntly	1	2	1:2 = 0.5	5
Lugton	2	3	2:3 = 0.67	3, 7
Leamington	4	3	4:3 = 1.33	1, 4, 6, 9
Morrinsville	4	3	4:3 = 1.33	2, 4, 7, 9
Matamata	5	4	5:4 = 1.25	2, 4, 5, 6, 9
Ngaruawahia	2	2	2:2 = 1	3, 8
Otorohanga	1	2	1:2 = 0.5	8
Paeroa	1	2	1:2 = 0.5	8
Ruakura	3	2	3:2 = 1.5	2, 4, 7
Taupiri	1	2	1:2 = 0.5	2
Te Aroha	2	3	2:3 = 0.67	4, 7
Te Awamutu	6	3	6:3 = 2	1, 2, 5, 6, 8, 9
Te Kuiti	2	2	2:2 = 1	1, 8
Thames	3	3	3:3 = 1	2, 4, 6
Taumarunui	1	4	1:4 = 0.25	6
Te Rapa	5	2	5:2 = 2.5	1, 3, 3, 5, 9

Table 1: Waikato Autumn Interclub men's entries, team:court ratios and division allocation

PHASE 1:
Applying the procedure given in Rasmussen & Trick, 2008 we produced the generic 6-team mirrored double round-robin draw shown in Table 2. Here the letters A—F denote the six teams, the + indicates a home match and the – an away match. We did not fill in the right-hand side of the Table as it is a "mirror" of the left-side (Weeks 1 to 5) with all the + and – signs being reversed. For the Waikato Autumn competition (a single round-robin event), only weeks 1 through 5 were played. For the Winter competition, weeks 6 through 10 will also be played and it is clear from this draw that any team has a maximum of three home or away games in successive weeks. For example, team D has three away games in succession from weeks 4 to 6 and team E has three home games in succession the same weeks. This result was shown in Rasmussen & Trick, 2008 to be the minimum break possible for any number of teams in a mirrored double round-robin tournament. If there are an odd number of teams, we need to insert a BYE team into one of the positions in the Table. Note that teams A-F, B-C, and D-E are complementary in the sense that their home and away sequences are reversed; for example, whenever A is at home then F will be away and vice versa. We exploit this complementarity property in the next phase when allocating teams from clubs with high team:court ratio.

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
A	F+	C-	E+	B-	D+					
B	E+	F-	D-	A+	C-					
C	D-	A+	F+	E-	B+					
D	C+	E-	B+	F-	A-					
E	B-	D+	A-	C+	F+					
F	A-	B+	C-	D+	E-					

Table 2: The generic 6-team mirrored double round-robin draw which minimises breaks

PHASE 2:
Next, we identify any clubs with a team:court ratio of 2 or greater to ensure all their courts are fully utilised every week. For our case study, there were only two such clubs being Te Awamutu and Te Rapa, having values of 6:3 = 2 and 5:2= 2.5 respectively. Clubs having ratios greater than 2 such as Te Rapa would normally require them to borrow a neighbouring club's court as a home venue to maintain the minimum break schedule. For these capacity clubs, we allocate their positions in the generic draw as a priority, also arranging their teams to be selected in complementary pairs. This ensures that those clubs will not have more teams at home than courts available but that all courts are being used every week. Te Awamutu entered six teams in Divisions 1, 2, 5, 6, 8 and 9 so for example, we allocate TA1=A, TA2=F, TA3=B, TA4=C, TA5= D, TA6=E. This allocation means that their division 1 and 2 teams are never at home together and neither are their division 5 and 6, or division 8 and 9 teams. Other pairings are also possible but for capacity clubs, exploiting the complementarity is essential to maintain full court utilisation.

PHASE 3:
To ensure that clubs who have entered two or more teams have at least two of those teams playing at home together every week, we allocate those teams the same identifier from the generic draw. As a precautionary measure to avoid scheduling more matches than available courts for smaller clubs (less than 3 courts available), those clubs entering 3 or more teams with a team:court ratio greater than 1 should have two of those teams paired complementarily. As an example, Morrinsville entered 4 teams and have 3 courts available, so allocating those teams to positions A and F in the generic draw ensures two teams are always at home. This precautionary measure is only necessary for those clubs with less than 3 courts though, as the generic six-team draw allocates exactly 3 home courts per week for each round.

3. RESULTS

Applying the three phases described in the Methods section to each of the nine divisions produces this possible allocation. Other allocations of the generic team positions A to F could be used but the break minimisation feature will always be maintained if the draw in Table 2 is followed.

Division 1	Allocation	Division 2	Allocation	Division 3	Allocation
Te Rapa 1	D	Morrinsville 1	C	Te Rapa 2	D
Te Kuiti 1	B	Matamata 1	E	Te Rapa 3	E
Leamington 1	F	Ruakura 1	A	Ngaruawahia 1	C
Te Awamutu 1	A	Taupiri	F	Cambridge 1	A
Hamilton 1	E	Te Awamutu 2	B	Lugton 1	F
Bye	C	Thames 1	D	Hamilton 2	B

Division 4	Allocation	Division 5	Allocation	Division 6	Allocation
Te Aroha 1	A	Te Rapa 4	E	Matamata 4	C
Morrinsville 2	C	Cambridge 2	A	Te Awamutu 4	D
Matamata 2	D	Matamata 3	D	Taumarunui	F
Ruakura 2	B	Huntly	F	Leamington 3	A
Leamington 2	F	Te Awamutu 3	C	Thames 3	B
Thames 2	E	Hamilton 3	B	Hamilton 4	E

Division 7	Allocation	Division 8	Allocation	Division 9	Allocation
Lugton 2	C	Paeroa	A	Te Rapa 5	B
Te Aroha 2	A	Cambridge 4	F	Morrinsville 4	E
Morrinsville 3	D	Ngaruawahia 2	C	Matamata 5	D
Cambridge 3	F	Otorohanga	D	Leamington 4	A
Ruakura 3	E	Te Kuiti 2	B	Te Awamutu 6	F
Hamilton 5	B	Te Awamutu 5	E	Hamilton 6	C

For ease of comparison with the automated draw produced by MySquash, a visual representation of both draws from Excel is shown in Figure 1, with colours to denote divisions. There are several distinguishing features of the draw produced using the minimum break methodology. Starting with the capacity clubs:

- Te Awamutu. All three courts are fully utilised every week without requiring the sequence of three home games for Division 6 in weeks 3 to 5.
- Te Rapa. With the additional requirement of requiring one overflow court in week 4, it is possible to bypass the MySquash sequence of three away games in succession for Division 5.

In the manual draw we see that the following desirable features are now included

- The Ngaruawahia, Te Kuiti and Te Aroha clubs have their two teams allocated at home on the same weeks.
- Otorohanga, Paeroa, Taupiri and Ruakura Div 2 no longer have a sequence of three home (or away) games.

Mens Court Allocations - Autumn Interclub 2026

Rounds:

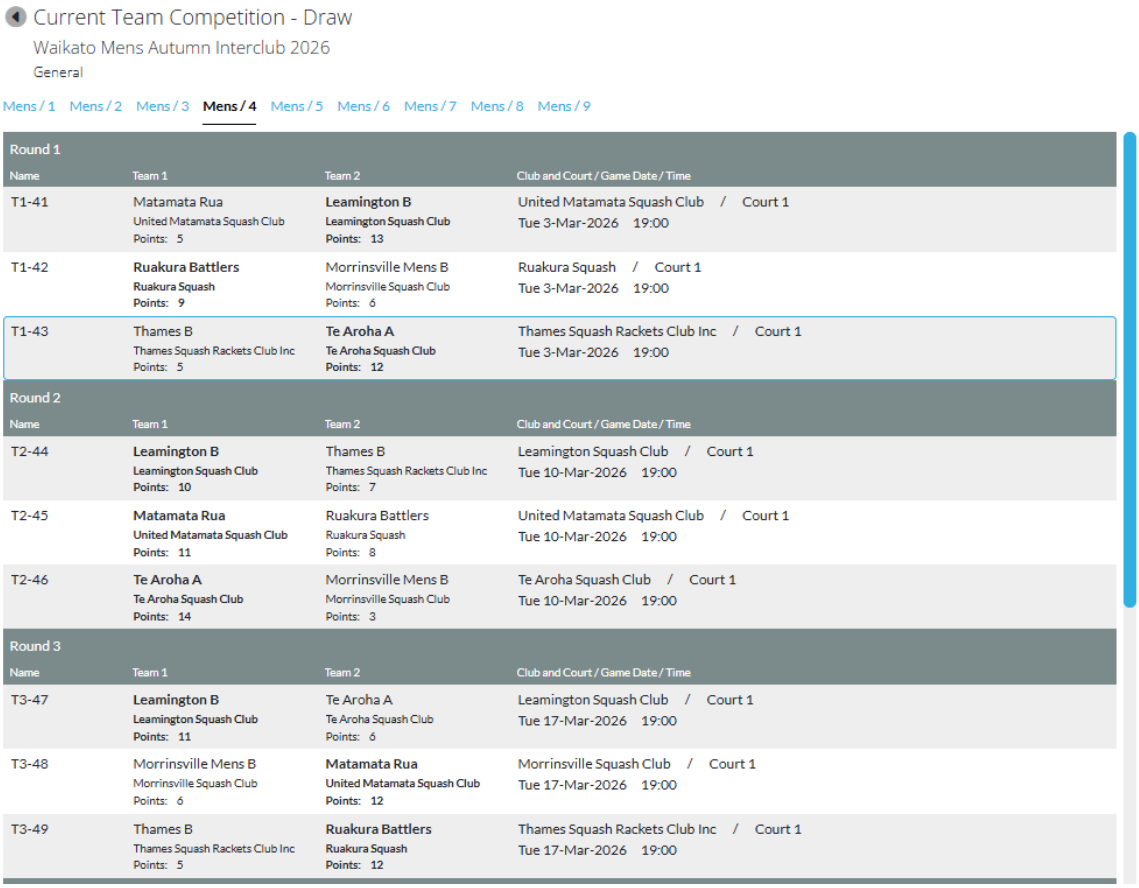
1	2	3	4	5		1	2	3	4	5
3-Mar	10-Mar	17-Mar	24-Mar	31-Mar		3-Mar	10-Mar	17-Mar	24-Mar	31-Mar
CAMB 1	CAMB 1	CAMB 1	CAMB 1	CAMB 1		CAMB 1	CAMB 1	CAMB 1	CAMB 1	CAMB 1
CAMB 2	CAMB 2	CAMB 2	CAMB 2	CAMB 2		CAMB 2	CAMB 2	CAMB 2	CAMB 2	CAMB 2
CAMB 3	CAMB 3	CAMB 3	CAMB 3	CAMB 3		CAMB 3	CAMB 3	CAMB 3	CAMB 3	CAMB 3
CAMB 4	CAMB 4	CAMB 4	CAMB 4	CAMB 4		CAMB 4	CAMB 4	CAMB 4	CAMB 4	CAMB 4
HAM 1	HAM 1	HAM 1	HAM 1	HAM 1		HAM 1	HAM 1	HAM 1	HAM 1	HAM 1
HAM 2	HAM 2	HAM 2	HAM 2	HAM 2		HAM 2	HAM 2	HAM 2	HAM 2	HAM 2
HAM 3	HAM 3	HAM 3	HAM 3	HAM 3		HAM 3	HAM 3	HAM 3	HAM 3	HAM 3
HAM 4	HAM 4	HAM 4	HAM 4	HAM 4		HAM 4	HAM 4	HAM 4	HAM 4	HAM 4
HAM 5	HAM 5	HAM 5	HAM 5	HAM 5		HAM 5	HAM 5	HAM 5	HAM 5	HAM 5
HAM 6	HAM 6	HAM 6	HAM 6	HAM 6		HAM 6	HAM 6	HAM 6	HAM 6	HAM 6
HUNT 1	HUNT 1	HUNT 1	HUNT 1	HUNT 1		HUNT 1	HUNT 1	HUNT 1	HUNT 1	HUNT 1
HUNT 2	HUNT 2	HUNT 2	HUNT 2	HUNT 2		HUNT 2	HUNT 2	HUNT 2	HUNT 2	HUNT 2
LUGT 1	LUGT 1	LUGT 1	LUGT 1	LUGT 1		LUGT 1	LUGT 1	LUGT 1	LUGT 1	LUGT 1
LUGT 2	LUGT 2	LUGT 2	LUGT 2	LUGT 2		LUGT 2	LUGT 2	LUGT 2	LUGT 2	LUGT 2
LUGT 3	LUGT 3	LUGT 3	LUGT 3	LUGT 3		LUGT 3	LUGT 3	LUGT 3	LUGT 3	LUGT 3
LEAM 1	LEAM 1	LEAM 1	LEAM 1	LEAM 1		LEAM 1	LEAM 1	LEAM 1	LEAM 1	LEAM 1
LEAM 2	LEAM 2	LEAM 2	LEAM 2	LEAM 2		LEAM 2	LEAM 2	LEAM 2	LEAM 2	LEAM 2
LEAM 3	LEAM 3	LEAM 3	LEAM 3	LEAM 3		LEAM 3	LEAM 3	LEAM 3	LEAM 3	LEAM 3
MORR 1	MORR 1	MORR 1	MORR 1	MORR 1		MORR 1	MORR 1	MORR 1	MORR 1	MORR 1
MORR 2	MORR 2	MORR 2	MORR 2	MORR 2		MORR 2	MORR 2	MORR 2	MORR 2	MORR 2
MORR 3	MORR 3	MORR 3	MORR 3	MORR 3		MORR 3	MORR 3	MORR 3	MORR 3	MORR 3
MATA 1	MATA 1	MATA 1	MATA 1	MATA 1		MATA 1	MATA 1	MATA 1	MATA 1	MATA 1
MATA 2	MATA 2	MATA 2	MATA 2	MATA 2		MATA 2	MATA 2	MATA 2	MATA 2	MATA 2
MATA 3	MATA 3	MATA 3	MATA 3	MATA 3		MATA 3	MATA 3	MATA 3	MATA 3	MATA 3
MATA 4	MATA 4	MATA 4	MATA 4	MATA 4		MATA 4	MATA 4	MATA 4	MATA 4	MATA 4
NGA 1	NGA 1	NGA 1	NGA 1	NGA 1		NGA 1	NGA 1	NGA 1	NGA 1	NGA 1
NGA 2	NGA 2	NGA 2	NGA 2	NGA 2		NGA 2	NGA 2	NGA 1	NGA 2	NGA 2
OTO 1	OTO 1	OTO 1	OTO 1	OTO 1		OTO 1	OTO 1	OTO 1	OTO 1	OTO 1
OTO 2	OTO 2	OTO 2	OTO 2	OTO 2		OTO 2	OTO 2	OTO 2	OTO 2	OTO 2
PAE 1	PAE 1	PAE 1	PAE 1	PAE 1		PAE 1	PAE 1	PAE 1	PAE 1	PAE 1
PAE 2	PAE 2	PAE 2	PAE 2	PAE 2		PAE 2	PAE 2	PAE 2	PAE 2	PAE 2
RUAK 1	RUAK 1	RUAK 1	RUAK 1	RUAK 1		RUAK 1	RUAK 1	RUAK 1	RUAK 1	RUAK 1
RUAK 2	RUAK 2	RUAK 2	RUAK 2	RUAK 2		RUAK 2	RUAK 2	RUAK 2	RUAK 2	RUAK 2
TAUP 1	TAUP 1	TAUP 1	TAUP 1	TAUP 1		TAUP 1	TAUP 1	TAUP 1	TAUP 1	TAUP 1
TAUP 2	TAUP 2	TAUP 2	TAUP 2	TAUP 2		TAUP 2	TAUP 2	TAUP 2	TAUP 2	TAUP 2
TARO 1	TARO 1	TARO 1	TARO 1	TARO 1		TARO 1	TARO 1	TARO 1	TARO 1	TARO 1
TARO 2	TARO 2	TARO 2	TARO 2	TARO 2		TARO 2	TARO 2	TARO 2	TARO 2	TARO 2
TARO 3	TARO 3	TARO 3	TARO 3	TARO 3		TARO 3	TARO 3	TARO 3	TARO 3	TARO 3
TAWA 1	TAWA 1	TAWA 1	TAWA 1	TAWA 1		TAWA 1	TAWA 1	TAWA 1	TAWA 1	TAWA 1
TAWA 2	TAWA 2	TAWA 2	TAWA 2	TAWA 2		TAWA 2	TAWA 2	TAWA 2	TAWA 2	TAWA 2
TAWA 3	TAWA 3	TAWA 3	TAWA 3	TAWA 3		TAWA 3	TAWA 3	TAWA 3	TAWA 3	TAWA 3
TKUI 1	TKUI 1	TKUI 1	TKUI 1	TKUI 1		TKUI 1	TKUI 1	TKUI 1	TKUI 1	TKUI 1
TKUI 2	TKUI 2	TKUI 2	TKUI 2	TKUI 2		TKUI 2	TKUI 2	TKUI 2	TKUI 2	TKUI 2
THAM 1	THAM 1	THAM 1	THAM 1	THAM 1		THAM 1	THAM 1	THAM 1	THAM 1	THAM 1
THAM 2	THAM 2	THAM 2	THAM 2	THAM 2		THAM 2	THAM 2	THAM 2	THAM 2	THAM 2
THAM 3	THAM 3	THAM 3	THAM 3	THAM 3		THAM 3	THAM 3	THAM 3	THAM 3	THAM 3
TAUM 1	TAUM 1	TAUM 1	TAUM 1	TAUM 1		TAUM 1	TAUM 1	TAUM 1	TAUM 1	TAUM 1
TAUM 2	TAUM 2	TAUM 2	TAUM 2	TAUM 2		TAUM 2	TAUM 2	TAUM 2	TAUM 2	TAUM 2
TAUM 3	TAUM 3	TAUM 3	TAUM 3	TAUM 3		TAUM 3	TAUM 3	TAUM 3	TAUM 3	TAUM 3
TAUM 4	TAUM 4	TAUM 4	TAUM 4	TAUM 4		TAUM 4	TAUM 4	TAUM 4	TAUM 4	TAUM 4
TRAP 1	TRAP 1	TRAP 1	TRAP 1	TRAP 1		TRAP 1	TRAP 1	TRAP 1	TRAP 1	TRAP 1
TRAP 2	TRAP 2	TRAP 2	TRAP 2	TRAP 2		TRAP 2	TRAP 2	TRAP 2	TRAP 2	TRAP 2
1	2	3	4	5		1	2	3	4	5
3-Mar	10-Mar	17-Mar	24-Mar	31-Mar		3-Mar	10-Mar	17-Mar	24-Mar	31-Mar

Div 1
Div 2
Div 3
Div 4
Div 5
Div 6
Div 7
Div 8
Div 9

Div 1
Div 2
Div 3
Div 4
Div 5
Div 6
Div 7
Div 8
Div 9

Figure 1: A comparison of the automated MySquash draw (left) and our manual method (right)

Figure 2 shows a screenshot of the first three rounds of the Division 4 draw produced by MySquash. The scores for each match have also been recorded. The notation used in MySquash is T1, T2,… for Rounds 1, 2,… with the number following being the match identifier. There was a total of 130 matches played over the five rounds and nine divisions, remembering that one division had a bye round each week ($9\times15-5=135-5=130$).



Current Team Competition - Draw

Waikato Mens Autumn Interclub 2026

General

Mens / 1 Mens / 2 Mens / 3 **Mens / 4** Mens / 5 Mens / 6 Mens / 7 Mens / 8 Mens / 9

Round 1

Name	Team 1	Team 2	Club and Court / Game Date / Time
T1-41	Matamata Rua United Matamata Squash Club Points: 5	**Leamington B** **Leamington Squash Club** **Points: 13**	United Matamata Squash Club / Court 1 Tue 3-Mar-2026 19:00
T1-42	**Ruakura Battlers** **Ruakura Squash** **Points: 9**	Morrinsville Mens B Morrinsville Squash Club Points: 6	Ruakura Squash / Court 1 Tue 3-Mar-2026 19:00
T1-43	Thames B Thames Squash Rackets Club Inc Points: 5	**Te Aroha A** **Te Aroha Squash Club** **Points: 12**	Thames Squash Rackets Club Inc / Court 1 Tue 3-Mar-2026 19:00

Round 2

Name	Team 1	Team 2	Club and Court / Game Date / Time
T2-44	**Leamington B** **Leamington Squash Club** **Points: 10**	Thames B Thames Squash Rackets Club Inc Points: 7	Leamington Squash Club / Court 1 Tue 10-Mar-2026 19:00
T2-45	**Matamata Rua** **United Matamata Squash Club** **Points: 11**	Ruakura Battlers Ruakura Squash Points: 8	United Matamata Squash Club / Court 1 Tue 10-Mar-2026 19:00
T2-46	**Te Aroha A** **Te Aroha Squash Club** **Points: 14**	Morrinsville Mens B Morrinsville Squash Club Points: 3	Te Aroha Squash Club / Court 1 Tue 10-Mar-2026 19:00

Round 3

Name	Team 1	Team 2	Club and Court / Game Date / Time
T3-47	**Leamington B** **Leamington Squash Club** **Points: 11**	Te Aroha A Te Aroha Squash Club Points: 6	Leamington Squash Club / Court 1 Tue 17-Mar-2026 19:00
T3-48	Morrinsville Mens B Morrinsville Squash Club Points: 6	**Matamata Rua** **United Matamata Squash Club** **Points: 12**	Morrinsville Squash Club / Court 1 Tue 17-Mar-2026 19:00
T3-49	Thames B Thames Squash Rackets Club Inc Points: 5	**Ruakura Battlers** **Ruakura Squash** **Points: 12**	Thames Squash Rackets Club Inc / Court 1 Tue 17-Mar-2026 19:00

Figure 2: The draw produced by MySquash for Division 4, showing the home venues and results for the first 3 weeks

4. DISCUSSION

Looking at some of the other districts in New Zealand, similar issues to our case study might arise although the travel issue may not be so problematic as here in the Waikato. We are an extremely geographically spread-out region with the farthest clubs being more than 270 km apart. For example, the Auckland district is also currently running a competition with ten men's and six women's divisions (six-team divisions). However, because of the population density of Auckland and the plethora of clubs, travel time between the farthest of these is probably no more than 1.5 hours compared to the 3.5 hours required in the Waikato. The Wellington district is also running a mixed competition with ten divisions, but their divisions only consist of four teams with a mirrored double round-robin format. This competition will also take 6 weeks to complete and there would be some travel issues, with club spread requiring travel time of around 2 hours.

Turning to the South Island, Canterbury ran a summer interclub competition with six men's and three women's divisions, each consisting of 6 teams. Although Canterbury is also a large geographic region, most of the clubs in this competition are located relatively close to Christchurch (the main city) so travel was probably not a factor in designing the draws. This competition was also unusual in that it ran for 7 weeks, with no indication of how the last two weeks were scheduled once the initial round-robin of 5 weeks was completed. Otago also ran a mixed competition with eight divisions of 6 teams with four divisions playing on Tuesdays and the other four on Wednesdays. There was no finals night, so these ran for a total of five weeks and only

involved clubs located close to the region's main city Dunedin. In contrast, Southland district has divided their clubs into East and West regions and given their players the choice of playing in those regions only, or to participate in a district-wide event. This may be an effective compromise to the travel time problem, as their 19 clubs are spread over an area even larger than Waikato's with more difficult road access.

We have presented only one possible allocation of teams to the generic schedule which was guided by the desire to a) minimise the number of breaks for each team, b) reward the capacity clubs by fully utilising their courts and c) allow clubs with multiple entries to have their teams playing at home together. It would be an interesting exercise to investigate whether an AI tool could assist in arranging the allocation so that all these goals were achieved for every club.

5. CONCLUSIONS

By choosing one possible allocation of the generic six-team draw for the Waikato 2026 Autumn Men's Interclub competition, we have shown that it is possible to capture most of the desirable features of a balanced draw whilst maintaining a minimum break schedule. These features are not produced by the automated MySquash draw and would probably require many iterations to replicate. Although each iteration only takes a few seconds to complete, the number of possible court allocations means a complete search is impossible.

The break minimisation procedure applied here may not seem of much importance in a single round-robin competition because each team has only two or three away games for a six-team division. However, it becomes more relevant for the mirrored double round-robin competition taking place in Winter when travel conditions become more difficult, and of further importance still for those districts running competitions with larger divisions. The district organisers should consider whether the small-time investment involved in allocating teams to the minimal break schedule (which also simultaneously allows for full court utilisation and club home teams playing together) outweighs the ease of generating an unsatisfactory draw automatically.

Acknowledgements

We wish to thank Squash N.Z. for allowing use and reproduction of the interclub draws produced by MySquash software, and access to the interclub district information for New Zealand.

References

Post, G., & Woeginger, G. J. (2006). Sports tournaments, home–away assignments, and the break minimization problem *Discrete Optimization, 3 (2),*165-173.

Rasmussen, R. V. & Trick, M. A. (2008). Round robin scheduling – a survey. *European Journal of Operational Research, 188*, 617-636.

Rosa, A. & Wallis, W. D. (1982). Premature sets of 1-factors or how not to schedule round-robin tournaments *Discrete Applied Mathematics, 4*, 291-297.

de Werra, D. (1981) Scheduling in sports. In P. Hansen (Ed.), Studies on Graphs and Discrete Programming, North-Holland, Amsterdam, pp. 381-395.

Squash New Zealand (n.d). MySquash Interclub Module. Retrieved from www.squash.org.nz

World Squash https://www.worldsquash.sport/qualification-system/. Retrieved April 4th, 2026.

List Management Optimisation in the Australian Football League

Hamish Murray[a,d], Paul Corry[a], Elliott Carr[a], Chris Drovandi[a]

[a] *Queensland University of Technology*
[d] *Corresponding author: Hamish.murray@hdr.qut.edu.au*

Abstract
List management as a mathematical concept is an application of operations research with aims to assist in planning longer-term decisions around the construction of a team, with the aims of improving the success of the team over this period based on specified objectives and within constraints. For the Australian Football League, as with many other professional leagues across the world, this involves a range of decisions in salary cap management, drafting strategy, free agency and trade, against other teams with the same broad set of goals. Given these constraints and equalisation present in the league, teams must navigate the peaks and troughs in performance likely to present themselves throughout the cycle. In developing this model, it aims to provide some assistance in how decisions relate to one another, how these may impact longer term and how optimising performance in target years may be set up in decisions made in earlier seasons. The optimisation methodology involves mean-variance optimisation where expected player performance by season is balanced with the relative risk involved in particularly drafting unproven players who may never achieve regular football at AFL level. Results showing the relative consistency of expected performance given different draft inputs and the importance of initial players to upcoming seasons will be shown. An extension to the model showing how updating for performance after one season impacts future performance through a lookahead model will also be demonstrated.

Keywords: Operations research, optimisation, list management, Australian Football

Submitted to MATHSPORT 2026

COMPARATIVE ANALYSIS OF DEFENSIVE PRESSURE IN MEN'S AND WOMEN'S AUSTRALIAN FOOTBALL

Steven Azzopardi [a,c,d], Minh Huynh [b], Jacqueline Tran [a], David Carey [a]

[a] *La Trobe University, Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services and Sport, Melbourne, Victoria, Australia*
[b] *Monash University, Department of Econometrics and Business Statistics, Clayton, Victoria, Australia*
[c] *Essendon Football Club, Melbourne*
[d] *Corresponding author: s.azzopardi@latrobe.edu.au*

Abstract
Defensive pressure is a concept used in team invasion sports to assess aspects of a team's match performance (Sullivan et al., 2014a; Sullivan et al., 2014b). More defensive pressure has been shown to be associated with poorer offensive skill execution (Browne et al., 2019; González-García, 2024; Link et al., 2016; Parrington et al., 2013). A quantification of defensive pressure in Australian football (AF), known as 'pressure acts' (*Official AFL Glossary*, 2019), was introduced by the Australian Football League's (AFL and AFLW) official data provider, Champion Data, to measure the level of defensive pressure applied to the ball carrier at the point of disposal in AF matches. This provided further contextual detail to the data collected in professional men's and women's matches. Pressure acts are a six-level ordinal variable, with levels: set, none, corralling, chasing, closing and physical. Previous studies in men's AF have found a correlation between increased defensive pressure on the ball carrier and decreased disposal efficiency (Browne et al., 2019; Browne et al., 2022; Parrington et al., 2013; Robertson et al., 2019; Walker et al., 2024). However, there are limited findings into the effects of defensive pressure in women's AF. This study aimed to compare expected pressure (Azzopardi et al., 2022) in elite women's and men's AF, and to investigate differences when adjusting or not adjusting for the context of disposals. Secondary aims were to test for changes in the distributions of pressure over time in each competition, and to estimate the effect of different amount of pressure on disposal efficiency in each competition.

Keywords: defensive pressure, team defence, invasion sport

Acknowledgements
We wish to thank Champion Data and their team of data capturers for the collection of the match data used for this study.

References
Azzopardi, S., Carey, D., Huynh, M., & Manson, B. (2022). Understanding pressure metrics and their importance in Australian football. Australasian Conference on Mathematics and Computers in Sport 2022,

Browne, P. R., Sweeting, A. J., Davids, K., & Robertson, S. (2019). Prevalence of interactions and influence of performance constraints on kick outcomes across Australian Football tiers: Implications for representative practice designs. *Hum Mov Sci*, *66*, 621-630.

Browne, P. R., Sweeting, A. J., & Robertson, S. (2022). Modelling the Influence of Task Constraints on Goal Kicking Performance in Australian Rules Football. *Sports Med Open*, *8*(1), 13.

González-García, I. (2024). Analysis of Game Variables and Influence on Defensive Success in Elite Handball. *Physical Education and Sports: Studies and Research*, *3*(2), 77-95.

Link, D., Lang, S., & Seidenschwarz, P. (2016). Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS One*, *11*(12), e0168768.

Official AFL Glossary. (2019). Champion Data. https://www.championdata.com/glossary/afl/]

Parrington, L., Ball, K., & MacMahon, C. (2013). Game-based analysis of handballing in Australian Football. *International Journal of Performance Analysis in Sport*, *13*(3), 759-772.

Robertson, S., Spencer, B., Back, N., & Farrow, D. (2019). A rule induction framework for the determination of representative learning design in skilled performance. *J Sports Sci*, *37*(11), 1280-1285.

Sullivan, C., Bilsborough, J. C., Cianciosi, M., Hocking, J., Cordy, J., & Coutts, A. J. (2014a). Match score affects activity profile and skill performance in professional Australian Football players. *J Sci Med Sport*, *17*(3), 326-331.

Sullivan, C., Bilsborough, J. C., Cianciosi, M., Hocking, J., Cordy, J. T., & Coutts, A. J. (2014b). Factors affecting match performance in professional Australian football. *Int J Sports Physiol Perform*, *9*(3), 561-566.

Walker, A., Dwyer, D. B., Young, C. M., & Bruce, L. (2024). Factors that affect handball execution in Australian Football. *J Sports Sci*, 1-6.

Approaches to Measuring and Modelling Player Development in Australian Rules Football

Dave Matteo [ad], David L. Carey [a], Joshua D. Ruddy [b] and Matthew C. Varley [a]
[a] *Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services and Sport, Latrobe University, Melbourne, Australia*
[b] *Essendon Football Club, Melbourne, Australia*
[d] *Corresponding author: dmatteo30@gmail.com*

Abstract
Evaluating player performance and development in Australian Rules Football (AF) has heavily relied on subjective judgments from coaches and scouts (MacMahon et al., 2019). With increasingly rich match-event data now available, there is an opportunity to assess performance and growth through objective, reproducible methods (Davis et al., 2024). This study develops a statistical framework for quantifying player development across multiple skill domains while accounting for contextual factors such as age, competition level, and playing position. First, this framework evaluates the meta-metric properties (Franks et al., 2016) of commonly used in-game AF performance metrics to identify which variables meaningfully discriminate between player performances and which remain stable across seasons. This provides an evidence-based foundation for selecting metrics that reliably measure player performance. Second, variables are combined using dimensionality-reduction techniques to derive composite indicators representing distinct skill domains and player role characteristics, such as ball distribution quality or defensive effectiveness. These composite indicators can then be tracked longitudinally to evaluate a player's historical development over time. Third, developmental baselines are estimated using time series models that account for age and competition level, allowing each player's observed trajectory to be compared with an expected trajectory. For each player, deviation from these expected curves can be used both to characterize prior development and predict future growth. By integrating metric validation, latent skill construction, and longitudinal modelling, this framework provides a data-driven approach to player evaluation and development in AF, with practical applications for recruitment, coaching, and talent development.

Keywords: Player Performance, Player Development, Player Prediction

Acknowledgements
We wish to thank Steven Azzopardi for his expertise in Australian Football Data and to Essendon Football Club for access to player data from Champion Data

References
Davis, J., Bransen, L., Devos, L., Jaspers, A., Meert, W., Robberechts, P., Van Haaren, J., & Van Roy, M. (2024). Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned. *Machine Learning*, *113*(9), 6977-7010.

Franks, A. M., D'Amour, A., Cervone, D., & Bornn, L. (2016). Meta-analytics: tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports*, *12*(4), 151-165.

MacMahon, C., Bailey, A., Croser, M., & Weissensteiner, J. (2019). Exploring the skill of recruiting in the Australian Football League. *International Journal of Sports Science & Coaching*, *14*(1), 72-81. https://doi.org/10.1177/1747954118809775

Submitted to MATHSPORT 2026

SAVE OUR SCORES: OFFSETTING AFL STRENGTH OF SCHEDULE IMBALANCE WITH POINTS RECALIBRATION

Anthony Bedford, Paul Hardwicke

University of the Sunshine Coast
abedford@usc.edu.au

Abstract
A level playing field is fundamental to competitive fairness in professional sport; however, the Australian Football League (AFL) operates a partial round robin format that introduces significant strength of schedule (SoS) disparity. This research addresses the perceived fairness imbalance by proposing a recalibrated premiership points system designed to diminish distortion caused by repeat match-ups. Under the current system, six repeat opponents are assigned based on a "weighted rule" that can fail to produce neutral outcomes, evidenced by a mean absolute deviation exceeding an acceptable benchmark of two matches.
This study analyses a 13-season data collection period (DCP) between 2012 and 2025. The "Veracity" model reduces the value of a win in a repeat match-up from four to three points, while maintaining four points for single match-ups. Retrospective application of this model across the DCP resulted in 32% of ladder positions changing. Notably, the 2025 season analysis reveals that the Veracity model delivers a 12% shallower points differential gradient. The research finds that while the current system favours commercially popular "marquee" repeat match-ups, it compromises ladder integrity. The Veracity model offers a viable mathematical intervention to preserve commercial interests while ensuring that final ladder rankings are biased-corrected indicators of true team quality.

Keywords: fairness, ladder, match-up, premiership points, strength of schedule

1. INTRODUCTION

Australian Football League (AFL) clubs compete for premierships. Critical to this pursuit is climbing the league ladder to qualify for finals participation. A level playing field over the course of a football season is considered a basic requirement for competitive fairness; however, fairness comprises competing objectives, such as playing against traditional rivals and competing based on strength. This research will address a perceived fairness imbalance in the AFL's scheduling format by proposing to reconfigure its premiership points allocation.
The AFL schedule has been unbalanced since 1987, with teams meeting some opponents once and others twice per season in a partial round robin format.
Since 2023, the 18 AFL teams have played 23 matches per season, each requiring six repeat match-ups. This creates "strength of schedule" (SoS) disparity, due to the disproportionate strength of opponents that teams play twice. This can make end-of-season ladders biased indicators of true team quality and distort competitive balance metrics (Lenten, 2010).
Scheduling repeat opponents provides some teams with softer assignments than others. A "free hit" against a weaker opponent can make a significant difference in end-of-season ladder position. For example, in season 2025, Adelaide won 75% of its 12 repeat match-ups and finished the season a game clear on top of the ladder. This position may have overstated its true strength, as evidenced by it failing to advance in the finals series, confirming Clarke's (1998) position that ladder positions can misrepresent teams' performances. This was consistent with an independent assessment of Adelaide's "fixture difficulty" at the start of the season, based on repeat opponent assignments, ranking it 15th out of 18 (Stats Insider, 2024).
AFL scheduling advantages some teams over others, which compromises ladder integrity. This can be improved upon by recalibrating premiership points awarded for repeat match-ups.
This research project proposes an alternative model to allocate AFL premiership points for repeat match-ups. It aims to quantify SoS bias, demonstrate how this bias can be diminished, and show how end-of-season ladder rankings can change as a result. This will involve developing a point-scoring model that distinguishes between single match-ups ("SoS singles") and repeat match-up ("SoS doubles") between opponents.

2. METHODS

SCHEDULING BACKGROUND

Teams advance on the AFL ladder by winning premiership points. Four points are awarded for a win and two points for a draw (though less than one percent of matches result in a draw). In the post-2022 18-team league, teams play 23 matches for a potential 92 points. Ladder rankings determine whether or not teams qualify for finals participation. Generally, the higher a team finishes on the ladder, the better their finals chances.

From 1994 to 2025, teams finishing in the top eight ladder positions contested finals series. The 2026 wildcard finals system will include ten teams, classifying them into the top four, positions five and six, and positions seven to ten. This will ensure that more teams are engaged for longer in the season to compete for finals places, as proposed by Lenten and Winchester (2010), and enable more teams to compete in finals, consistent with the AFL's objective of broad finals participation across the league (De Hollander et al., 2025; Jakee et al., 2020).

Constraints can influence scheduling bias. They are foundations of AFL scheduling that dictate who, where, and when each team will play. Some constraints are opaque under secrecy requirements, such as agreements with venues (Jakee et al., 2022). "Soft" constraints act as rules of thumb that can be set aside when necessary; "hard" constraints should invalidate the schedule if broken (Kyngäs et al., 2017). Examples include avoiding scheduling teams for consecutive home or away matches (soft), and that no team can play with fewer than five recovery days between matches (hard).

The AFL's weighted rule is a predictive fairness strategy that assumes past performance correlates with future strength. It is intended to maintain competitive fairness by matching opponents of similar strength. The intent can be compromised by marquee games such as rivalry match-ups (Davies, 2013). This research seeks to provide insurance against this.

At season's end (prior to finals), teams are classified into the top six, middle six, and bottom six ladder rungs for the purpose of allocating SoS doubles opponents for the following season. The method of allocating these opponents is subject to a set of constraints that collectively comprise the weighted rule (Smith, 2024).

DATA COMPILATION

Data from 13 seasons were compiled in Excel to test the research question and hypothesis. Seasons 2012–2019 and 2021–2025 were included as they represent a stable period where 18 teams competed in seasons of either 22 or 23 matches. These comprise the data collection period (DCP). Season 2020 was excluded as only 17 rounds were played due to the COVID-19 pandemic; it included no SoS doubles.

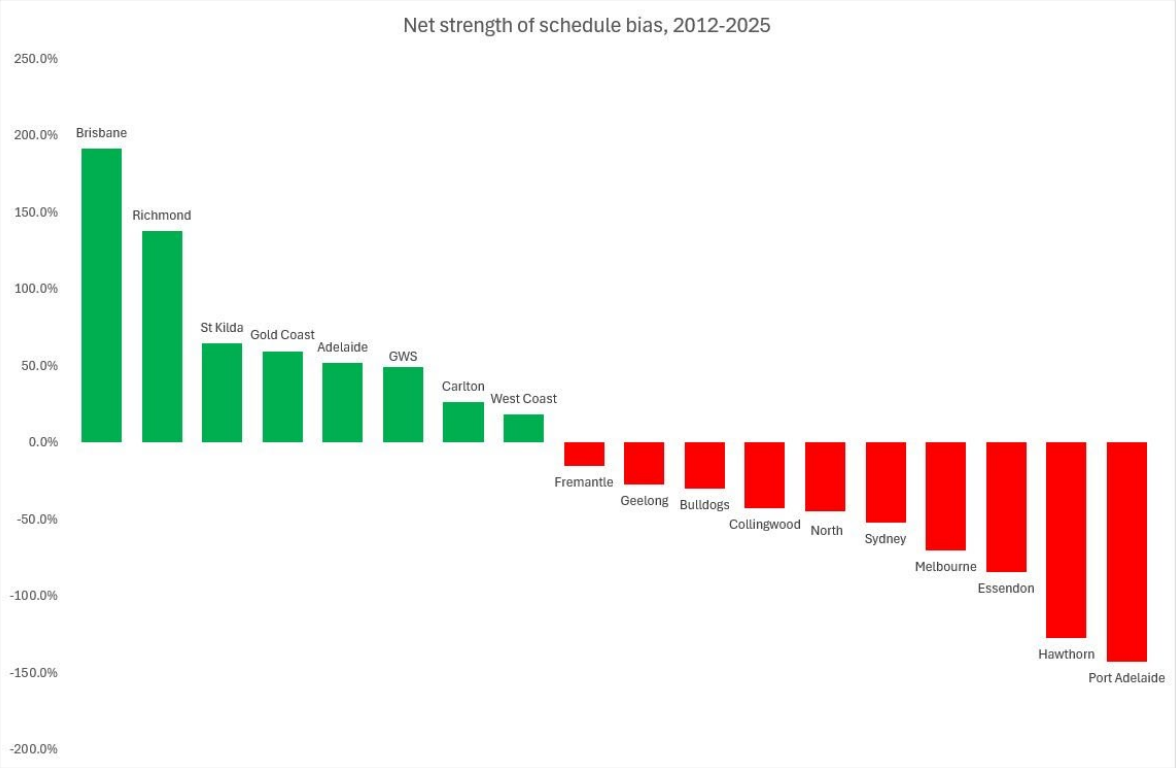
ESTABLISHING BIAS

In a balanced season, teams would win the same percentage of SoS doubles as they do SoS singles. A perfectly middling performer, such as the Western Bulldogs in 2017, would win (and lose) half of its SoS singles and half of its SoS doubles matches.

Net SoS bias by team was assessed to identify which teams were the biggest winners and losers over the DCP. This was achieved by comparing each team's performance in SoS singles versus their SoS doubles matches. It does not factor in the scoreboard magnitude of wins or losses.

Chart 1 below shows the variance between teams' SoS performances and indicates net SoS bias over the DCP, ranking teams from most advantaged to most disadvantaged. Where these differences approached zero, their schedule was deemed to be balanced with their respective strength. The larger the discrepancy, the greater the imbalance of their repeat assignments and the greater their schedule volatility. Chart 1 highlights the advantages held by Brisbane and Richmond by factors of three and two respectively over the next nearest rival, St Kilda. Brisbane's cumulative bias was 1.9 matches or 7.7 premiership points, equivalent to 0.6 points per season. The five teams in the band between Carlton and the Western Bulldogs moved by fewer than 30 percentage points in either direction, suggesting they had the fairest SoS double schedules. Port Adelaide and Hawthorn were the most disadvantaged teams. The distance in SoS bias between Brisbane and Port Adelaide was 3.3 points over the DCP, or 0.3 points per season.

Chart 1: Net "strength of schedule" bias by team



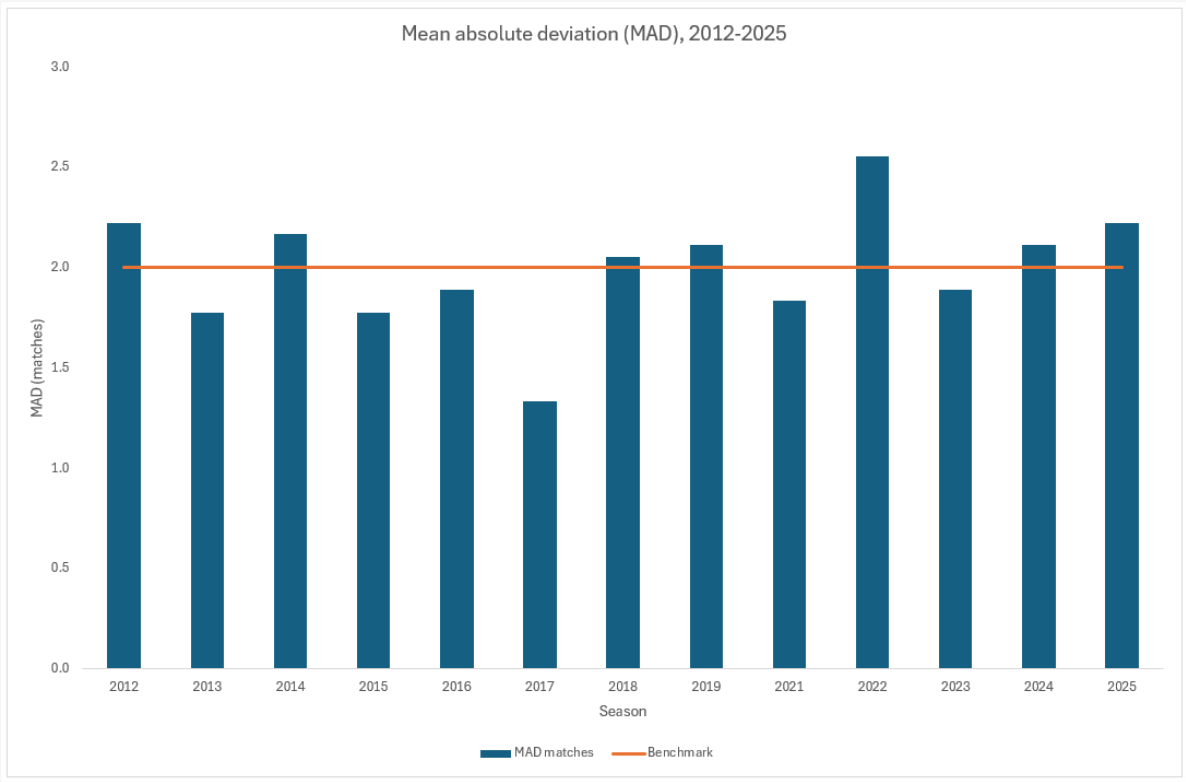
APPROACH

The methodology leaned into the work of Lenten (2010), by quantifying the bias inherent in the schedule between SoS singles and SoS doubles, so that premiership points could be adjusted within-season. The adjustment was retrospectively applied to the DCP, and notional changes to ladder positions observed. Lenten separated single match-ups, double match-ups, and hypothetical "notional" games needed for a dual round robin schedule. He calculated the average quality of opponents that teams' faced once and those they faced twice, inferred results for the notional matches, and produced adjusted win ratios. He applied these win ratios to the (then) 16 AFL teams over 12 seasons (1997–2008), and observed changes at a maximum of just under one win per team per season.

This study used mean absolute deviation (MAD) to quantify scheduling bias inherent in the AFL's current schedule format. This metric provided a transparent measure of horizontal equity, representing the average amount of "unfairness" experienced by a representative team compared to the league mean. The 24 points available from the six SoS doubles matches comprise the potential bias in the schedule. The MAD was calculated for each team within each season, and then averaged for that season. Each team's total SoS doubles wins (W) were assessed against a benchmark (B) of the number of wins in a perfectly balanced draw (half the SoS doubles matches played; i.e., B = 5.0 for 2012-2022 and B = 6.0 for 2023-2025).

Chart 2 below shows that in a typical season, the MAD from a six-win neutral benchmark is approximately two matches, or eight premiership points. It demonstrates that the premiership point bias was inherent to the 12 SoS doubles matches. The Y-axis represents the MAD in terms of matches. A higher MAD indicates a season where the fixture was less equitable, meaning some teams had significantly easier or harder paths than others. The orange line at 2.0 matches (or eight premiership points) represents the benchmark for an acceptable level of schedule deviation. Chart 2 illustrates that 2017 was the most equitable season in the DCP, with a MAD well below 1.5. Season 2022 was the most unfair or unbalanced season, spiking above 2.5.

Chart 2: SoS double wins mean absolute deviation (MAD)



MODEL PROPOSAL

Lenten and Winchester (2010) addressed imbalance the ladder placings strength by awarding bonus premiership points based on winning margins. This research proposes to minimise the effects of SoS bias on teams' ladder rankings by using a points-recalibration logic to absorb the distortion created by SoS doubles.

The MAD bias is worth eight premiership points per season. When divided by the 12 SoS doubles matches, this is equivalent to two-thirds of a point per match. This means that 3.3 points is the unbiased value of a SoS double match. For simplicity, and longevity as the competition prepares to grow in 2028 and beyond, a value of three premiership points is appropriate.

Four premiership points are maintained as the value for a SoS single win (called a four-pointer), with three points for a SoS double win (a three-pointer), and two points for a drawn SoS single. A drawn SoS double is set at 1.5 points for the model's mathematical integrity. Further research into whether this passes the pub test, and what to do about it if not, will be addressed separately. For brevity, the proposal is hereafter referred to as the "Veracity" model.

Under this model, the six repeat match-ups (now worth 18 points, six fewer than six matches at four points) represent 22.5% of the 80 available premiership points. This is a reduction of 3.6 percentage points from the 26.1% imbalance in the actual schedule, and compresses potential bias by six premiership points, effectively bringing all teams 1.5 matches closer together.

Table 1 below demonstrates the numbers of SoS singles and SoS doubles played in seasons 2012-2022 and 2023-2025. The current (post-2022) period includes six matches for 24 points that double up on single round robin matches. More than a quarter (26.1%) of the 92 available premiership points are derived from this weighted component. It shows the variance in points available in seasons of 22 matches (2012 to 2022) and 23 matches (2023 to 2025), for the standard (actual) and Veracity models.

Table 1: Points comparison under standard and Veracity models

	2012 to 2022				2023 to 2025			
	Matches	% total	Points	% total	Matches	% total	Points	% total
Actual season								
Single round robin @ 4 points	17	77.3%	68	77.3%	17	73.9%	68	73.9%
Additional matches @ 4 points	5	22.7%	20	22.7%	6	26.1%	24	26.1%
Total	**22**		**88**		**23**		**92**	
SoS singles @ 4 points	12	54.5%	48	54.5%	11	47.8%	44	47.8%
SoS doubles @ 4 points	10	45.5%	40	45.5%	12	52.2%	48	52.2%
Total	**22**		**88**		**23**		**92**	
Veracity model								
Single round robin @ 3 & 4 points	17	77.3%	63	80.8%	17	73.9%	62	77.5%
Additional matches @ 3 points	5	22.7%	15	19.2%	6	26.1%	18	22.5%
Total	**22**		**78**		**23**		**80**	
SoS singles @ 4 points	12	54.5%	48	61.5%	11	47.8%	44	55.0%
SoS doubles @ 3 points	10	45.5%	30	38.5%	12	52.2%	36	45.0%
Total	**22**		**78**		**23**		**80**	
Variance								
Single round robin	0	0	-5	3.5%	0	0	-6	3.6%
Additional matches	0	0	-5	-3.5%	0	0	-6	-3.6%
Total	**0**		**-10**	**0%**	**0**		**-12**	**0%**
SoS singles	0	0%	0	7.0%	0	0%	0	7.2%
SoS doubles	0	0%	-10	-7.0%	0	0%	-12	-7.2%
Total	**0**		**-10**	**-11.4%**	**0**		**-12**	**-13.0%**

Match data were flagged with opponent pairs meeting either once (SoS singles) or twice (SoS doubles) in a season. Premiership points were applied separately under actual conditions and the Veracity model. League tables were constructed in 13 individual worksheets for each season. Teams were tagged as top eight finals qualifiers, wildcard finals qualifiers, and into weighted rule cohorts.

An example of how the Veracity model would impact ladder placings in season 2025 is provided below in Table 2. Fremantle would have finished fifth ahead of Greater Western Sydney, and Port Adelaide would have jumped from 13th to 11th position, bumping St Kilda into the bottom weighted rule cohort.

Table 2: Impact of Veracity model on ladder placings, season 2025

Season 2025	Veracity ladder											
Team	Played	Won	Lost	Drawn	For	Against	%	Singles won @ 4 pts	Doubles won @ 3 pts	Singles drawn @ 2 pts	Doubles drawn @ 1.5 pts	Premiership points
Adelaide	23	18	5	0	2278	1635	139.3%	36	27	0	0	63
Geelong	23	17	6	0	2425	1714	141.5%	40	21	0	0	61
Brisbane	23	16	6	1	2061	1804	114.2%	28	27	2	0	57
Collingwood	23	16	7	0	1991	1627	122.4%	32	24	0	0	56
Fremantle	23	16	7	0	1978	1815	109.0%	32	24	0	0	56
Greater Western Sydney	23	16	7	0	2114	1834	115.3%	28	27	0	0	55
Gold Coast	23	15	8	0	2173	1740	124.9%	32	21	0	0	53
Hawthorn	23	15	8	0	2045	1691	120.9%	32	21	0	0	53
Western Bulldogs	23	14	9	0	2493	1820	137.0%	24	24	0	0	48
Sydney	23	12	11	0	1845	1902	97.0%	16	24	0	0	40
Port Adelaide	23	9	14	0	1705	2136	79.8%	24	9	0	0	33
Carlton	23	9	14	0	1799	1861	96.7%	12	18	0	0	30
St Kilda	23	9	14	0	1839	2077	88.5%	12	18	0	0	30
Melbourne	23	7	16	0	1902	2038	93.3%	16	9	0	0	25
Essendon	23	6	17	0	1535	2209	69.5%	16	6	0	0	22
North Melbourne	23	5	17	1	1805	2365	76.3%	4	12	2	0	18
Richmond	23	5	18	0	1449	2197	66.0%	4	12	0	0	16
West Coast	23	1	22	0	1466	2438	60.1%	4	0	0	0	4

BENCHMARKING FAIRNESS

By simulating equitable points distribution over time, the research assessed the relative fairness of actual and proposed ladders by comparing their relative proximity to an equity benchmark.

Assuming SoS equity with all 18 teams equally likely to succeed, probabilities of teams achieving certain results (e.g., qualifying for the double chance) a certain number of times were calculated using binomial distribution B(n,p). Results were benchmarked as target distributions for ladder placings.

Probabilities (p) were calculated as 22% for a team to qualify for a double chance (finish in any of the top 4 places out of 18 teams); 44% to qualify for the finals (top 8 places); 56% to qualify for the wildcard finals (top 10 places); and 33% to finish in any particular weighted rule cohort (one of three groups of six places).

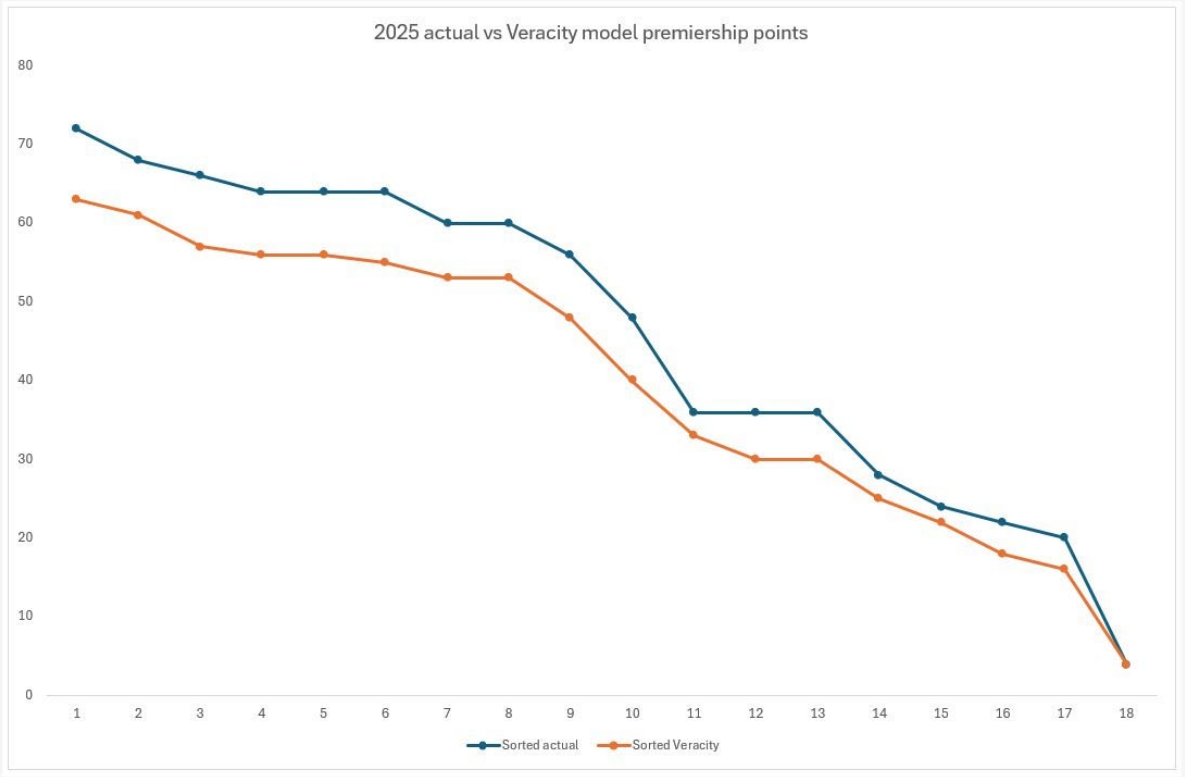
The results in Table 3 below apply the total team count (18) multiplied by the binomial probability of success. That is, the expected number of teams is equal to 18 x P(X = k) where P(X = k) is the binomial probability of exactly k successes in 13 trials (representing 13 AFL seasons).

Table 3: Binomial distribution probability of teams' representation in ladder sections

Frequency	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Double chance (top 4 finish)	0.7	2.5	4.4	4.6	3.3	1.7	0.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Standard finals (top 8 finish)	0.0	0.1	0.4	1.3	2.5	3.6	3.9	3.1	1.9	0.8	0.3	0.1	0.0	0.0
Wildcard finals (top 10 finish)	0.0	0.0	0.1	0.3	0.8	1.9	3.1	3.9	3.6	2.5	1.3	0.4	0.1	0.0

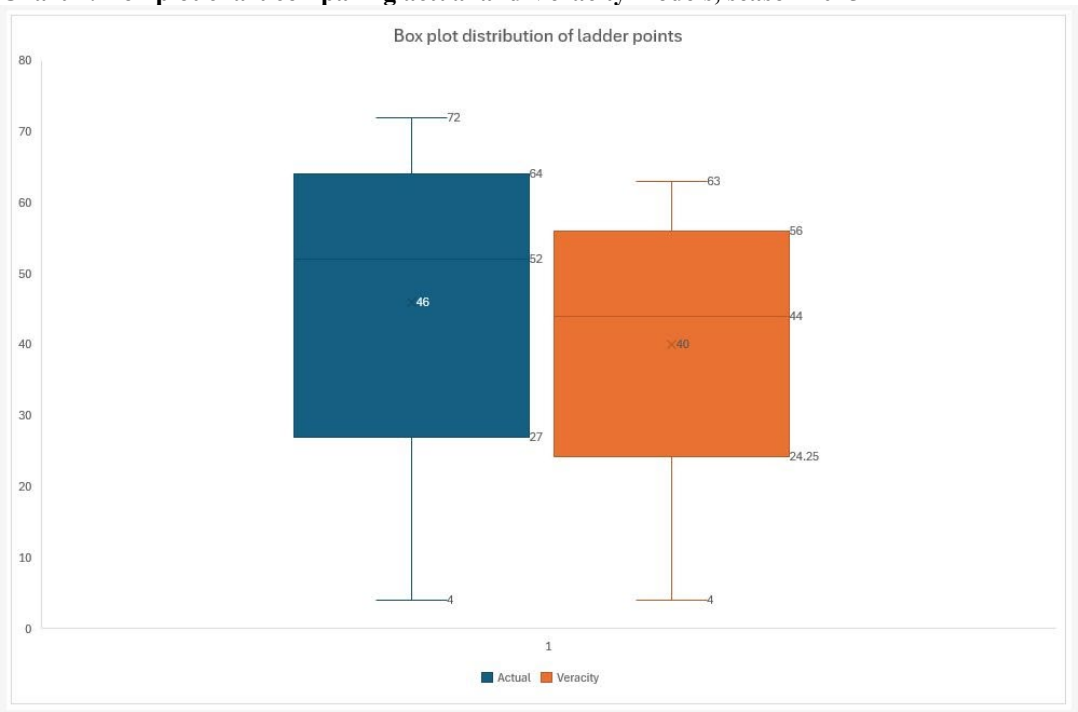
Chart 3 below shows that Veracity delivered a more competitive, even ladder. In the actual season, the gap between a top-four team and a mid-table team was steeper and more punishing than the Veracity model. The shallower slope in the season 2025 Veracity model indicates that a fairer schedule that can reduce the dominance of elite teams. This compressed slope shows that it successfully creates a closer, more balanced competition compared to the reality of the 2025 season, with a measurable reduction in structural dominance.

Chart 3: Actual and Veracity models points earned by team slope, season 2025



The box plot in Chart 4 below confirms the compression of premiership points in the Veracity model. The actual interquartile range (IQR) spans 37 points. The Veracity IQR spans 31.75 points. This highlights a shift in where the middle of the competition sits compared to the actual season. The middle 50% is more tightly grouped. Its shorter top whisker, reaching 63 instead of 72, indicates that a dominant outlier is statistically less likely than under actual conditions. The actual median is 52 points and Veracity median is 44 points, the latter being more central and an indicator of a more balanced season. The box plot illustrates that the Veracity model can compress points differential, improve equality, and find a fairer middle ground for the competition.

Chart 4: Box plot chart comparing actual and Veracity models, season 2025



3. RESULTS

By applying the Veracity model over the 13 seasons and 234 total ladder positions of the DCP, 76 changes (32%) would have occurred. The maximum number of ladder movements by a team in a single season was three places. Only once would a team have displaced another in the final eight. The Western Bulldogs would have lost eighth position to Carlton in season 2022.

Qualification for the wildcard top 10 finals was classified into the four aforementioned groups. Should this system have been in place across the DCP, 18 teams would have moved between the four groups, with the maximum number of movements in a season being one change. The mid-section of the wildcard ladder showed the most volatility, with 10th position being surrendered to 11th on four occasions, and ninth to 11th position twice. Under Veracity, six teams would have competed in finals that would not have done under the standard points allocation system. In 2012, Essendon would have displaced Carlton. In 2014, Collingwood would have bumped Adelaide. In 2017, St Kilda would have displaced the Western Bulldogs. In 2018, North Melbourne would have surrendered eighth position for 11th, with Port Adelaide moving to 9th and Essendon the beneficiary in 10th position. In 2021, St Kilda and Fremantle would have advanced at the expense of West Coast. In 2023, Essendon would have displaced Adelaide.

Qualification for the three weighted rule cohorts would have resulted in eight changes between the groups, with the maximum in a season being one change.

Each team's total absolute deviation reflected its cumulative drift from a neutral schedule over the DCP. The Western Bulldogs averaged 1.4 matches of deviation per year, meaning that their performance in SoS doubles aligned more closely with their SoS singles than any other team. Their low deviation scores meant their final ladder positions reliably reflected their strength rather than their schedules. North Melbourne averaged 2.6 matches of deviation per year, meaning that their final ladder positions were the most distorted by the uneven nature of the SoS doubles. Their scores demonstrated that the weighted rule does not produce neutral outcomes for all clubs, representing the weakest case from a competitive fairness perspective.

4. DISCUSSION

The 32% change to ladder placings over the DCP, and maximum movement of three places, aligns exceptionally well with the work of Lenten (2010). Out of 192 possible placings, 68 teams (35.4%) in Lenten's study finished in different positions, with a maximum ranking change of three places.

The disparity between North Melbourne (34.0) and the Western Bulldogs (18.0) in total absolute deviation illustrates the extent of longitudinal schedule inequality under the standard points system. It accurately reflected the performance of teams like the Bulldogs, but generated volatility for teams like the Kangaroos, whose ladder positions were influenced by MAD values 30% higher than the league average. This demonstrates that the weighted rule can fail the teams that are in most need of fairer treatment.

It is noteworthy that the Western Bulldogs are the team most accurately scheduled according to the MAD metrics, and yet they are the only team to be forced out of the final eight under a recalibrated model. This highlights a false positive in the current AFL ladder system.

Two rivalry matches per season are hard-coded into the schedule for eight teams: Adelaide/Port Adelaide, Brisbane/Gold Coast, Fremantle/West Coast, and Sydney/Greater Western Sydney. These matches present a potentially unfair points bonus for the traditionally stronger of each pairing. For example, Brisbane had a 21:7 head-to-head record with struggling rival Gold Coast in the DCP, winning 74% of the available premiership points. Brisbane also finished in the top six every season from 2019 to 2025, and contested the 2023, 2024, and 2025 grand finals. The Collingwood, Carlton, and Essendon triumvirate rivalries are not stated explicitly in constraints, but are implied and often feature twice per season. Strongest of the three, Collingwood has won the greatest share of these matches in the DCP, with both its rivals having middling results in recent seasons.

Bringing teams 1.5 matches closer together in the league ladder is significant. In 12 of the 13 DCP seasons, the difference between eighth and ninth position was less than six points.

The Veracity model would complement Davies' system to help to minimise biases associated with mismatched opponents. By applying the Davies system, Port Adelaide would play three (theoretically) stronger opponents from the middle group and only one from the bottom group; noting that a sixth SoS double must also be scheduled in the current 23-match season format. Similarly, St Kilda would benefit from a softer schedule.

The primary challenge for the Veracity model is the "pub test," that is, the public and commercial perception of changing a century-old scoring tradition. While the status quo is simple, its cost is structural SoS bias. The Veracity model compressed the 2025 points differential gradient by 12%, effectively bringing the league 1.5 matches closer together and reducing the "gift" of a soft draw.

Critics may argue that reducing repeat match-ups to three points may diminish the emotional stakes of high-profile or marquee games. However, the model maintains a regulated ceiling; winning both doubles against a rival still nets six points, a 50% increase over a single match-up. Although traditionalists may resist over-engineering, AFL history shows that fans adapt to frequent rule changes if ladder integrity is preserved.

To mitigate complexity, the transition should be supported by visual transparency tools. These include introducing a mobile app featuring "live" shadow ladders and match ratios to socialise the concept in season 2027, before the formal 2028 system launch. Reintroducing the match ratio (last used in the AFL in 1994) would allow stakeholders to identify a team's true quality as the season progresses, regardless of their mix of three- and four-pointers. Table 4 below illustrates the match ratio as a snapshot of season 2025, round six.

Table 4: Veracity model league table featuring match ratio

2025 ladder after round 6, ranked by match ratio, points, %	Played	Four-pointers (singles) played	Three-pointers (doubles) played	Byes	Maximum points	For	Against	Match ratio	Points earned	%
Gold Coast	6	3	3	1	20	710	428	85.0%	17	165.9%
Brisbane	6	2	4	1	20	539	510	85.0%	17	105.7%
Collingwood	6	4	2	1	22	541	396	81.8%	18	136.6%
Geelong	6	4	2	1	22	589	462	72.7%	16	127.5%
Adelaide	6	4	2	1	22	652	494	68.2%	15	132.0%
Hawthorn	6	4	2	1	22	533	492	68.2%	15	108.3%
Greater Western Sydney	6	4	2	1	22	521	387	63.6%	14	134.6%
Essendon	6	6	0	1	23	488	617	52.2%	12	79.1%
Western Bulldogs	6	2	4	1	20	571	519	50.0%	10	110.0%
St Kilda	6	2	4	1	20	532	588	50.0%	10	90.5%
Fremantle	6	3	3	1	21	542	518	47.6%	10	104.6%
Port Adelaide	6	4	2	1	22	523	533	45.5%	10	98.1%
Richmond	6	4	2	1	22	420	639	31.8%	7	65.7%
Carlton	6	2	4	1	20	524	429	30.0%	6	122.1%
Sydney	6	2	4	1	20	498	493	30.0%	6	101.0%
Melbourne	6	4	2	1	22	412	600	18.2%	4	68.7%
North Melbourne	6	1	5	1	19	512	704	15.8%	3	72.7%
West Coast	6	3	3	1	21	368	666	0.0%	0	55.3%

5. CONCLUSIONS

The current AFL scheduling framework, while commercially successful, operates under a persistent bias where a team's ladder position can be as much a product of its fixture as its performance. This research demonstrates that the Veracity model provides a robust, transparent mechanism to mitigate this disparity by recalibrating the value of repeat match-ups.

Literature suggests that while the AFL successfully maintains high levels of interest through marquee match-ups and outcome uncertainty, the bias created by the partial round robin fixture remains a significant challenge. This research demonstrates that recalibrating premiership points offers a viable pathway to diminishing this imbalance, provided that the model is robust enough to account for the inevitable criticism that would be directed towards the AFL following its release.

The Veracity model mathematically addresses scheduling inequity. Teams with SoS doubles against lower-ranked sides gain fewer points per win, effectively compressing the advantage of a soft draw.

Reducing the premiership points awarded for SoS repeat pairings increases the outcome uncertainty desired by the AFL and supports extending the race for final ladder positions to the latest possible round.

The 2025 season shown in Chart 2 has a MAD of approximately 2.2, above the 2.0 benchmark, indicating that the 2025 schedule had a high level of inequality. Though final eight displacement happened only once in the DCP, the fact that the Western Bulldogs would have lost their 2022 finals spot illustrates the high stakes of the SoS bias in the current model. Similar displacement in a final 10 system is evidently more likely.

By reducing the points awarded for SoS doubles from four to three, the model effectively compresses structural dominance by elite teams and reins in advantages gifted by soft draws. Applying the Veracity model over the DCP would have adjusted almost a third of ladder positions, ensuring that finals structures more accurately reflect true team quality.

While the shift to a dual-point system introduces new complexity for fans and stakeholders, alignment to a mathematical ideal represents a significant step towards a more level playing field.

Acknowledgements

We wish to thank the team at kaggle.com for access to historical AFL match data.

References

Booth, R. (2006). The Economic Development of the Australian Football League. In S. Szymanski & W. Andreff (Eds.), *Edward Elgar Publishing eBooks*. Edward Elgar Publishing.

Clarke, S.R. (1998). Evaluating the fairness and efficiency of the Australian Football League home and away schedule. *ASOR bulletin, 17*, 2–11.

Davies, C. (2013). Competition parity and the Australian Football League fixtures. *Sporting traditions*, 30, 61.

DeHollander, A., & Karwan, M. (2025). Improving strength of schedule metrics in sports scheduling. *Journal of Quantitative Analysis in Sports*.

Ferguson, P., & Lakhani, K. (2023). Consuming Contests: The Effect of Outcome Uncertainty on Spectator Attendance in the Australian Football League*. *Economic Record.*

Jakee, K., Kenneally, M., & Dineen, D. (2022). Scheduling slots and league objectives: An empirical analysis of Australia's AFL. *Quarterly Review of Economics and Finance, 84,* 359–371.

Kyngäs, N., Lilley, G., Salter, T., & Goossens, D. (2017). Scheduling the Australian Football League. *Journal of the Operational Research Society*, *68*(8), 973–982.

Lee, Y. H., Kim, Y., & Kim, S. (2019). Competitive balance with unbalanced schedules. *Journal of Quantitative Analysis in Sports*, *15*(3), 239–260.

Lenor, S., Lenten, L. J. A., & McKenzie, J. (2016). Rivalry Effects and Unbalanced Schedule Optimisation in the Australian Football League. *Review of Industrial Organization*, *49*(1), 43–69.

Lenten, L. J.A. (2010). The extent to which unbalanced schedules cause distortions in sports league tables. *Economic Modelling*, *28*(1), 451–458.

Lenten, L., & Winchester, N. (2010). Optimal Bonus Points in the Australian Football League. *Economic Papers: A journal of applied economics and policy*, *29*(4), 407–420.

Reade, J., & Van Ours, J. (2024). Consumer Perceptions Matter: A Case Study of an Anomaly in English Football. *Jahrbücher für Nationalökonomie und Statistik*, 244, 605–629.

Scarf, P., Khare, A., & Alotaibi, N. (2021). On skill and chance in sport. *Ima Journal of Management Mathematics*.

Sharma, A. (2024). Impact of three-point rule change on competitive balance in football: a synthetic control method approach. *Applied Economics*, 57, 9264–9274.

Smith, M. (2024, November 14). How the fixture is made: Five-day breaks, Super Saturday, weighted rule. *Australian Football League.*

Smith, M. (2025, February 16). *Rep footy, marquee games on cards in AFL's push to 10 million. Australian Football League.*

TOWARDS HYDRATION SENSING: A DIMENSION REDUCTION FRAMEWORK FOR IMPEDANCE SPECTROSCOPY DATA

Alex Lap Cheung Liu [a,b,d] , Kerrie Mengersen [a], Paula Sanz-Leon [b], Paul Pao-Yen Wu [a]

[a] *Queensland University of Technology, Centre of Data Science*
[b] *WearOptimo*
[d] *Corresponding author: alex.liu@hdr.qut.edu.au*

Abstract
Impedance spectroscopy is an emerging, non-invasive, wearable biosensing technology for real-time monitoring of biological signals such as the hydration status of athletes in training and competition (Armstrong, L. E. (2007)). WearOptimo's non-invasive microwearable applies alternating currents between pairs of electrodes across a wide range of excitation frequencies and measures the opposition to current flow, called impedance (Abasi, S. et al. (2022)). A key challenge in analysing data from the device is the high dimensionality of impedance measurements, which often requires dimensionality-reduction techniques to improve computational scalability, support exploratory analysis, and extract meaningful features. Often in bioimpedance frequency sweeps, the measured responses are strongly correlated across frequencies, enabling the data to be represented by a smaller set of variables that account for most of the information in the measurements (Grimnes, S., & Martinsen, Ø. G. (2015)). However, there can be nonlinear coupling between resistance and reactance over different frequencies, which are factors frequently overlooked in existing literature. This study develops and evaluates the suitability of several dimensionality-reduction approaches, including Principal Component Analysis (PCA), complex PCA, functional PCA, and related techniques. The serial and sparse nature of impedance measurements from low to high frequencies introduces additional challenge for dimension reduction. Potentially, two-dimensional splines that smooth over resistance and reactance over the frequency spectrum can be used in functional PCA to develop lower dimensional and explainable representations of impedance spectroscopy data. Demonstrated on electrical test circuit dataset, the evaluation considers theoretical assumptions, interpretability, and robustness in capturing the underlying impedance dynamics. Ongoing work aims to extend these methods to more complex media such as saline solution, along with modelling tools and model selection, contributing to the development of interpretable and informative models for next-generation wearable physiological monitoring systems.

Keywords: Impedance spectroscopy, dimension reduction, principal component analysis, microwearables

Acknowledgements
We wish to thank WearOptimo for facilitating the experiments and supplying the resulting data. This project is made possible by CSIRO's Next Generation Artificial Intelligence Graduates Program (GA221787) funded by the Australian Government.

References
Abasi, S., Aggas, J. R., Garayar-Leyva, G. G., Walther, B. K., & Guiseppi-Elie, A. (2022). Bioelectrical impedance spectroscopy for monitoring mammalian cells and tissues under different frequency domains: A review. ACS Measurement Science Au, 2(6), 495–516.
Armstrong, L. E. (2007). Assessing hydration status: The elusive gold standard. Journal of the American College of Nutrition, 26(sup5), 575S-584S.
Grimnes, S., & Martinsen, Ø. G. (2015). Bioimpedance and bioelectricity basics (3d edition). Elsevier/Academic Press.

Submitted to MATHSPORT 2026

A Data-Informed Framework to Classify and Analyse Pacing Profiles in Competitive Swimming

Mackenzie Parker [a,b,c,f], Geoffrey M. Minett [d], Jack Fahey-Gilmour [e], Koji Honda [e], Paul Pao-Yen Wu [a,b]

[a] *School of Mathematical Sciences, Faculty of Science, Queensland University of Technology, Brisbane, Australia.*
[b] *Centre for Data Science, Queensland University of Technology (QUT), Brisbane, Queensland, Australia.*
[c] *Data61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia.*
[d] *School of Exercise and Nutrition Sciences, Faculty of Health, Queensland University of Technology (QUT), Brisbane, Australia.*
[e] *Queensland Academy of Sport (QAS), Queensland Government, the State of Queensland.*
[f] *Corresponding author: mackenzie.parker@hdr.qut.edu.au*

Abstract
This research presents a novel framework for identifying and interpreting pacing profiles in elite swimming. Pacing is the deliberate distribution of energy resources during athletic performance to optimise outcomes and mitigate fatigue (Foster et al., 2003). In swimming, pacing involves managing energy expenditure and race strategy to complete the event in the shortest time possible (Abbiss & Laursen, 2008; Foster et al., 1994). Pacing has become widely accepted as important determinants of success (Skorski et al., 2013), and optimal strategies vary between individuals. High-resolution velocity data from international men's 100m ($n = 164$), women's 100m ($n = 150$), men's 200m ($n = 175$), and women's 200m ($n = 173$) freestyle events were analysed using functional data analysis for the ability to capture intra-lap variation in velocity, address autocorrelation in time series data, and account for variability in skill execution. Race data was then clustered using k-means, hierarchical, and Gaussian mixture modelling methods to learn quantitative definitions and parameters for pacing profiles. Outputs from multiple clustering algorithms were combined using Bayesian model averaging (Forbes et al., 2023), reducing clustering algorithm biases and accounting for pacing profile classification uncertainty by informing profile definitions with multiple latent structures. These analysis findings were interpreted as conventional pacing profiles – such as positive, even, and parabolic (Abbiss & Laursen, 2008; McGibbon et al., 2018) – by contextualising their observed characteristics, providing simple communication of analysis results to coaches and athletes and synergy with current swimming practices. Race time distributions were compared across clusters using non-parametric statistics, revealing that pacing strategy was associated with performance outcomes, with significant race time differences in the women's events. The framework demonstrated robustness across clustering algorithms and is adaptable to different swimming events and data resolutions. These findings provide coaches and practitioners with a reproducible, data-informed approach to characterising pacing strategies and optimising performance in elite swimming.

Keywords: Functional Data Analysis, Unsupervised Learning, Elite Sport Performance, Performance Analysis, Sports Data Science.

Acknowledgements
This project was made possible by CSIRO's Next Generation Artificial Intelligence Graduates Program (GA221787) funded by the Australian Government. It was made in collaboration with the Centre for Data Science at QUT, the Queensland Academy of Sport, and Swimming Australia.

References
Abbiss, C. R., & Laursen, P. B. (2008). Describing and understanding pacing strategies during athletic competition. *Sports medicine*, *38*, 239-252.

Forbes, O., Santos-Fernandez, E., Wu, P. P.-Y., Xie, H.-B., Schwenn, P. E., Lagopoulos, J., Mills, L., Sacks, D. D., Hermens, D. F., & Mengersen, K. (2023). clusterBMA: Bayesian model averaging for clustering. *PLoS One*, *18*(8), e0288000. https://doi.org/10.1371/journal.pone.0288000

Foster, C., De Koning, J. J., Hettinga, F., Lampen, J., La Clair, K. L., Dodge, C., Bobbert, M., & Porcari, J. P. (2003). Pattern of Energy Expenditure during Simulated Competition. *Medicine & Science in Sports & Exercise*, *35*(5), 826-831. https://journals.lww.com/acsm-msse/fulltext/2003/05000/pattern_of_energy_expenditure_during_simulated.16.aspx

Foster, C., Schrager, M., Snyder, A. C., & Thompson, N. N. (1994). Pacing strategy and athletic performance. *Sports medicine*, *17*(2), 77-85.

McGibbon, K. E., Pyne, D., Shephard, M., & Thompson, K. (2018). Pacing in swimming: A systematic review. *Sports medicine*, *48*, 1621-1633.

Skorski, S., Faude, O., Caviezel, S., & Meyer, T. (2013). Reproducibility of Pacing Profiles in Elite Swimmers. *International journal of sports physiology and performance*. https://doi.org/10.1123/IJSPP.2012-0258

MATHEMATICAL MODELING OF ADAPTIVE PAYMENT RAIL SWITCHING IN HIGH-DEMAND SPORTS EVENTS

Ajoudanian, Shohreh[1], Aboutalebi, Hamid Reza [2], Saremi, Shahrzad [3], Nooraei Abadeh, Maryam [4] , Mirzaei, Mansooreh [3], Shibl, Rania [5], Bedford, Anthony [3], Mealy, Erica [3], Khaled Aldakheel, Shaden[6]

[1,2] Department of Computer Engineering, Na.C., Islamic Azad University, Najafabad, Iran
[3] School of Science, Technology and Engineering, University of the Sunshine Coast, Brisbane, QLD 4502, Australia
[4] Department of Computer Engineering, Arv. C., Islamic Azad University, Abadan, Iran
[5] Faculty of Science and Engineering, Southern Cross University, Gold Coast, Australia
[6] School of Art and Design, Princess Nourah Bint Abdulrahman University, Riyadh 11564, Saudi Arabia

Abstract. Major international events such as the Olympic Games, the FIFA World Cup, and other large-scale national competitions that attract crowds place significant pressure on payment systems. These events are characterized by sudden surges in demand, numerous transactions happening simultaneously, and strict availability requirements. If payments fail or are delayed, it can result not only in direct financial losses but also in serious social and reputational issues. This paper introduces DPOS (Dynamic Payment Optimization System), a flexible policy engine designed to improve payment system performance during high-demand events. These events often cause rapid increases in transaction volume that can overwhelm payment systems, leading to economic setbacks and damage to reputation if failures occur. DPOS addresses this by automatically switching between various payment methods, such as card networks and digital wallets, while optimizing for transactional risk, cost, and latency. It utilizes real-time system health monitoring and dynamic risk-budgeting to minimize losses during demand spikes. The framework includes components for signal collection, decision-making, secure execution of switches, and activity logging. Testing with real-world fraud data and network scenarios shows that DPOS significantly improves authorization success rates and reduces costs compared to traditional approaches, thereby enhancing operational resilience and regulatory compliance in high-throughput payment environments.

Keywords: DPOS, Payment systems, High-demand sporting events, Transaction optimization, Operational resilience

INTRODUCTION

The rapid growth in digital payment volumes and the expanding variety of payment methods have fundamentally changed the structure of modern payment systems. These systems are increasingly adopting multi-rail architectures. In such systems, a single transaction can be routed across multiple payment rails, including card networks, instant bank transfers, QR payments, and digital wallets, depending on availability, performance, and risk factors (Su et al., 2024). This architectural change is no longer just about boosting efficiency; it is now crucial for systems to operate reliably in unstable environments.

Adaptive multi-rail payment architectures are particularly crucial during major sporting events such as the FIFA World Cup, the Olympic Games, and large national competitions with high attendance. In these settings, payment systems must handle sudden demand spikes, process numerous transactions simultaneously, and meet strict availability standards. In such scenarios, payment failures or high end-to-end latency not only cause immediate financial losses for merchants and service providers but also damage their reputation and trust in digital payment systems.

Although there have been significant advances in instant payment rails and wallet-based solutions, issues related to risk and security persist, particularly when connectivity is poor or the device is offline (Howell, 2025). Offline payments carry risks like double-spending and delayed settlement, so robust methods are needed to verify and reconcile payments once the connection is restored. Recent research indicates that integrating reliable hardware primitives, such as Trusted Platform Modules (TPMs), Trusted Execution Environments (TEEs), and monotonic counters, can significantly enhance the security of offline payment systems (Reddy et

al., 2024). QR-based payment systems are popular because they are inexpensive to implement and compatible with a wide range of devices. However, they remain vulnerable to fraud techniques like code manipulation, replay attacks, and image spoofing. Therefore, lightweight cryptographic and authentication methods are essential (Okubo et al., 2024).

In addition to securing each payment rail, the process of selecting and switching rails presents a unique and largely unexplored challenge (Slivkins et al., 2024; Pacchiano et al., 2024). This paper emphasizes the need for a single, deployable, dynamic policy engine for multi-rail payment systems, a need not addressed by the many solutions proposed to date. We introduce DPOS, a flexible policy framework that:

• transforms selecting a payment rail into a multi-objective online optimization problem balancing risk, cost, and latency;

• utilizes multiple layers of operational and security signals to assess the rail's health in real time;

• enforces operational limits through clear penalties and adjustable risk budgets; and

• maintains regulatory accountability through modifiable policies, replayable decisions, and audit logs that record any tampering.

The main contribution of this work is not simply to propose a new payment rail or security primitive, but to demonstrate how it integrates adaptive decision-making, risk-aware switching, and auditability into a unified architecture suitable for large-scale, high-throughput payment environments.

Related Work

Research on the resilience of the Internet and service ecosystems consistently shows that major network disruptions can swiftly affect interconnected services, including payments, financial systems, and travel (Nurmi et al., 2025). These studies emphasize the importance of redundancy, adaptive routing, and graceful degradation to maintain service continuity under adverse conditions (Hashimoto et al., 2025). While these studies offer valuable insights into systemic vulnerabilities and resilience strategies, they mainly remain descriptive or infrastructural and lack operational decision models specifically designed for the unique constraints of payment systems.

To enhance security without overburdening computers, experts have proposed lightweight cryptographic schemes, dual-message QR encodings, and machine-learning-based forgery detection (Chou & Chen, 2024). While these innovations strengthen individual payment channels, they do not address how to dynamically select or integrate these channels into a comprehensive payment system.

Auditability and non-repudiation have become crucial requirements for financial systems. To maintain integrity and detect unauthorized modifications, tamper-evident logging systems based on hash chains, Merkle trees, and blockchain-inspired designs have been suggested (Morillo Reina et al., 2025). Hybrid on-chain/off-chain architectures, in which cryptographic commitments are stored on-chain and bulk data is kept off-chain, have proven effective for balancing scalability and verifiability (Yoon, 2025).

According to the systematic review in *Sensors*, NFC-based rails are vulnerable to relay, eavesdropping, and cloning attacks, and their security assessment must be considered when selecting rails. This finding indicates that a dynamic policy engine should not rely solely on performance metrics like latency and error rate but should also incorporate security signals into its decision-making process (Onumadu et al., 2024; Li et al., 2024).

The literature collectively establishes solid foundations across four areas: (i) infrastructure resilience, (ii) secure offline and alternative payment systems, (iii) tamper-evident auditing mechanisms, and (iv) restricted online decision-making. The main issue is that there is no single operational policy engine that can dynamically select payment rails while optimizing risk, cost, and latency, and meeting regulatory audit requirements. DPOS was specifically developed to address this gap.

Problem Formulation and System Architecture

The goal of DPOS is to select the optimal payment method for each incoming transaction in real time and online, taking into account operational, regulatory, and security requirements. This approach helps reduce overall financial losses. The available options include a limited set of payment methods such as card networks, instant account-to-account transfers, QR-based payments, and digital wallets that function both online and offline. From the policy engine's perspective, each transaction is treated as an atomic event, and the status of each payment method changes in response to observed signals. Offline-capable methods are assumed to include risk-control mechanisms such as trusted hardware (TPM/TEE), secure elements, or monotonic counters that prevent unlimited double-spending and enable settlement and reconciliation once connectivity is restored.

Multi-Objective Cost Function and Constraints

Choosing a payment rail is akin to solving a multi-objective optimization problem, where the system must balance three competing factors: risk, cost, and latency. These factors are normalized to a common scale to enable rapid decision-making. This can be achieved by normalizing monetary values or using standardized scores weighted by shadow prices set by the business. For a candidate rail (a) at decision time (t), the instant objective function is defined as:

$$\lambda\Omega_t(a_t) + \omega_{\hat{L}} L_t(a_t) + \omega_{\hat{C}} C_t(a_t) + \omega_{\hat{R}} R_t(a_t) = \min_{a_t \epsilon A} J_t(a_t)$$

where $R_t(a_t)$ denotes the expected loss associated with transactional risk (e.g., fraud, decline, or double-spending), $C_t(a_t)$ represents the direct and indirect monetary cost of using rail (a), and $L_t(a_t)$ captures the economic impact of end-to-end latency. The weights $\omega_{\hat{L}}$, $\omega_{\hat{C}}$, and $\omega_{\hat{R}}$ reflect business preferences or seasonal priorities, and reduce to a single measure when all components are expressed in monetary terms. The penalty term $\Omega_t(a_t)$ encodes violations or near-violations of operational constraints, scaled by the coefficient (λ). Latency-induced loss is represented either by quantifying delay through its effect on abandonment probability or by integrating standardized delay scores. In both cases, higher latency increases the objective value and reduces the likelihood that people will choose slow rails.

To prevent excessive accumulation of exposure, DPOS enforces a dynamic risk budget over a sliding window of length W. Let $\hat{L}_t$ denote the realized or expected risk loss at time t. The cumulative risk consumption over the window must satisfy:

$$\sum_{k=t-W+1}^{t} R_k\widehat{(a_k)} \leq B_t$$

where B_t stands for the acceptable risk budget, which may vary by rails, channels, or merchant segments. As cumulative consumption gets closer to B_t, the policy engine automatically moves to lower-risk rails, even if they cost more or take longer.

Switching Stability Constraint

Changing payment rails too frequently can degrade the user experience and complicate operations. To lessen this effect, DPOS adds a clear switching cost, which is defined as:

$$a_t = arg\min_{a \epsilon A}(J_t(a) + \gamma[a \neq a_{t-1}])$$

where a_{t-1} denotes the previously selected rail and γ is a configurable penalty coefficient. This term is added to the objective function, ensuring that rail changes occur only when the expected benefit outweighs the switching penalty.

Signal Processing and Metric Estimation

DPOS maintains three small estimators for each payment rail: a Risk Estimator, a Cost Estimator, and a Latency Estimator. These estimators constantly receive new information from multiple layers of operational and security signals.

Application-layer signals include metrics such as the frequency of authorization failures, API failures, timeouts, and response times. Transport and network-layer signals track factors such as round-trip time (RTT) (Güler et al., 2025), jitter, packet loss, DNS resolution failures, and connection resets. Contextual security signals involve elements specific to each rail, such as QR anti-forgery markers, NFC threat alerts, offline counter states, and policy acceptance rules. To balance responsiveness and stability, signals are normalized using weighted sliding windows. Simple confidence intervals or variance proxies measure the uncertainty of these estimates. These are incorporated into the objective function to enable more conservative decision-making when data is unclear.

The System Architecture

DPOS is designed as a modular, extensible system to support large-scale payment networks while ensuring transparency, resilience, and auditability. Its core parts work together to maintain operational efficiency and cryptographic security. The Signal Collector continuously gathers telemetry from payment terminals, gateways, and network probes, and can optionally perform external health checks on each network segment to evaluate overall system readiness. The Estimators convert these raw signals into real-time, parameterized estimates of risk, cost, and latency, enabling adaptive updates as conditions change. The Decision Arbiter uses multi-objective optimization to balance performance, stability, and compliance. It incorporates constraints, hysteresis, and risk budgets before issuing precise rail selection commands. The Switch Executor safely performs these switches, manages partial rail failures, and ensures transactional continuity. Lastly, the Tamper-Evident Log Store maintains an append-only ledger with cryptographic hashing, timestamping, and Merkle root verification. Bulk telemetry and transaction data are stored off-chain, while their cryptographic digests are anchored to a verifiable ledger, ensuring scalable, provable integrity. Through this system, DPOS dynamically chooses the best rail based on real-time conditions and risk exposure, without sacrificing transparency, compliance, or operational performance.

Dynamic Policy Engine and Degraded Operating Modes

DPOS is a comprehensive policy engine built to optimize payment routing under both normal and adverse network conditions. It employs a dynamic decision core that balances economic efficiency maximizing the transaction success rate (ASR) with cost per transaction (CPT) and latency (P95) while maintaining strict operational safety. Before any learning-based optimization, DPOS enforces a set of precedence rules: hard filters that immediately exclude non-compliant or SLA-violating rails; a rail kill switch that temporarily disables rails affected by large-scale outages; and a deterministic fallback sequence that ensures predictable behavior when multiple rails are unavailable. The system's auditability framework separates decision logic which uses aggregated, anonymized signals—from audit logs, which retain only derived evidence. It also enforces role-based access to proof bundles, minimizing personally identifiable information while retaining enough evidence for regulatory and forensic analysis. Evaluation involves two steps: offline replay of transaction sequences with realistic latency profiles, and canary-style validation of a synthetic traffic subset under feature flags. Statistical significance is assessed using Fisher's exact or Z-tests for rate-based metrics, and Mann-Whitney U tests for latency and cost distributions. In benchmarks against fixed-priority, round-robin, and late-switch baselines—each sharing the same limits, fees, and SLA thresholds—DPOS showed significant, simultaneous improvements in success rate, latency, and cost without increasing SLA violations or risk-budget overflows. It maintains near-zero switch rates due to hysteresis and switching penalties. Overall, the engine provides resilient, cost-effective routing that preserves user experience, operational predictability, and full auditability under both normal and degraded network conditions.

At its core, DPOS models payment rail selection as a constrained contextual bandit problem. At each decision time (t), the policy engine observes a context vector summarizing the current system state and selects a rail (a_t) from the available action set. For each candidate rail (a), the policy computes an augmented decision score:

$$S_t(a) = -\left(\omega_R \widehat{R_t}(a) + \omega_C \widehat{RC_t}(a) + \omega_L \widehat{L_t}(a)\right) + \beta_t(a)\sigma_t(a)$$

where $\sigma_t(a)$ is the multi-objective cost defined in Equation (1), $\beta_t(a)$ is an uncertainty or exploration bonus derived from estimation variance or confidence bounds, and $\beta_t(a)$ controls the exploration–exploitation trade-off. Unlike classical bandit formulations, exploration in DPOS is explicitly risk-aware. The exploration coefficient $\beta_t(a)$ is adaptively scaled based on the remaining risk budget:

$$B_{t+1} = \max\{0, B_t - \widehat{R_t}(a_t)\} + r_t$$

where B_t is the risk budget allocated for the current window. When the budget decreases, exploration slows, which pushes the policy toward safe, well-known paths. This mechanism prevents loss amplification under stress or during long-term degradation.

As shown in Table 1, DPOS uses three different stability mechanisms to keep the user experience and operational predictability.

Table 1. Decision Stability Mechanisms

Mechanism	Purpose	Effect on Policy
Hysteresis threshold	Avoid marginal switches	Filters noise-driven improvements
Minimum dwell time	Stabilize user experience	Prevents rapid oscillation
Switching cost (\gamma)	Penalize frequent changes	Requires meaningful gain to switch

Also, DPOS uses a set of strict safety rules that take precedence over optimization logic in certain situations before using the learning-based decision core (Table 2).

Table 2. Rule-Based Overrides in Extreme Conditions

Condition	Action	Rationale
Legal or compliance violation	Immediate rail exclusion	Non-negotiable requirement
Critical SLA breach	Rail kill-switch	Protect system-wide performance
Multi-rail outage	Deterministic fallback	Ensure continuity of service

When the network experiences partial or complete failure, DPOS enters a controlled-degraded mode. In this mode, rails that can work offline are turned on, and strict rules about limited risk as shown in Table 3.

Table 3. Operating Modes and Policy Behavior

Operating Mode	Active Rails	Risk Handling	Logging Strategy
Normal	All eligible rails	Full optimization	Central tamper-evident log
Degraded	Subset of rails	Conservative budgeting	Central + local buffering
Offline	Offline-capable rails only	Strict bounded risk	Local log → post-sync

Using weighted sliding windows, the estimator parameters and policy weights are continuously updated. Controlled adjustments to weighting coefficients allow for the incorporation of business-level preferences such as seasonal priorities, channel sensitivity (POS versus online), or campaign-specific goals. When empirical abandonment curves are available, latency monetization can be activated. If not, standardized latency scores are employed alongside shadow pricing. All updates are versioned to track changes.

Auditability and Proof Framework

In regulated payment systems, adaptive decision-making must be supported by strong guarantees of explainability, integrity, and non-repudiation. The framework relies on three key principles to ensure that DPOS can withstand internal audits, regulatory review, and investigations after an incident: (i) tamper-evident evidence creation, (ii) deterministic replay of decisions, and (iii) data minimization to safeguard privacy. The core idea is that each routing decision leaves a small, verifiable trace sufficient to reconstruct and verify the decision logic without requiring trust in the system operator or exposure to sensitive raw data.

Evidence Generation and Decision Replay

DPOS maintains a proof bundle for each transaction at the moment of decision. This bundle includes the policy context, a summary of the inputs, outputs, and constraint states used by the decision arbiter. When the routing decision is made, proof bundles are created simultaneously and added to a log that can only expand. Each proof bundle allows for deterministic replay. An auditor can recalculate the objective function, verify that the constraints were satisfied, and confirm that the selected rail was the optimal choice under the current policy by using the stored policy version and summarized inputs. This design prevents policy changes after the fact and ensures that operational actions can be traced.

Tamper-Evident Logging Architecture

The logging subsystem uses a tamper-evident log (TEL) constructed with cryptographic hash chaining and Merkle tree aggregation. Each record links to the hash of the previous entry, ensuring that data can only be added. Merkle roots summarizing log segments are generated regularly and cryptographically connected to a verifiable ledger or a trusted timestamping service.

DPOS uses a hybrid on-chain and off-chain design to balance scalability and verifiability: off-chain storage maintains operational metadata and bulk proof bundles, while on-chain (or ledger-anchored) storage records only cryptographic commitments like hashes and Merkle roots. This method offers strong guarantees of integrity without creating excessive storage or latency. The audit framework is based on clear claims that must be provable later. Each claim includes supporting evidence and a verification method.

Table 4. Audit Evidence Map for DPOS Decisions

Audit Claim	Evidence Artifact	Storage Location	Verification Method	Audit Purpose
Decision followed active policy	Policy ID and version signature	Proof bundle + log	Match policy hash with policy repository	Prevent post hoc policy changes
Inputs were unaltered	Input digest + Merkle path	Proof bundle + off-chain store	Merkle membership proof	Enable decision replay
Risk budget respected	Pre-/post-decision budget state	Proof bundle	Recompute cumulative risk	Prevent excessive exposure
SLA constraints enforced	SLA flags and metrics	Proof bundle	Threshold verification	Regulatory compliance
Switching penalty applied	(\gamma) and indicator term	Proof bundle	Re-evaluate objective function	User experience stability
Log integrity preserved	Hash chain + Merkle roots	Append-only log	Consistency and inclusion proofs	Detect deletion or tampering
Timestamp validity	Multi-source time attestation	Proof bundle	Signature verification	Prevent time rollback attacks
Offline mode bounded	Monotonic counter / token IDs	Local → central sync	Counter reconciliation	Double-spend prevention

DPOS defines a standard proof bundle schema to ensure more consistent evidence generation. The schema is designed to be concise, protect privacy, and be sufficient for independent verification.

Table 5. Proof Bundle Schema

Field	Description	Type / Format	Privacy Consideration
PolicyID, Version	Identifier of active policy	Text / semantic version	No PII
Context Slice	Channel, region, amount bucket	Categorical	No user identifiers
Inputs Digest	Hashed signals and metrics	JSON + hash	Aggregated statistics only
Decision	Selected rail and objective value	JSON	No sensitive data
Constraints State	Risk budget, SLA flags	JSON	Aggregated
Switch Cost	(\gamma) and indicator value	Numeric	—
Log Proofs	Hash chain and Merkle root	Hashes	Public-verifiable
Time Attestation	Multi-source timestamp	Binary / signature	No PII
Optional Signatures	System or third-party signature	Binary	Optional

Proof bundles never include raw telemetry or sensitive transaction data. Instead, only cryptographic summaries and combined metrics are stored, in accordance with the principles of data minimization and privacy-by-design.

Offline Evidence Handling and Reconciliation

In degraded or offline modes, DPOS creates local proof bundles and stores them securely in a local buffer. To prevent replay attacks or double-spending, each offline decision is linked to a monotonic counter or a cryptographically bound token. When connectivity is restored, the local logs are synchronized with the central tamper-evident log store. This process verifies counter continuity, validates the local hash chains, and integrates local Merkle roots into the global log, ensuring that offline activity is fully auditable and remains consistent with online operations.

Evaluation, Experimental Setup, and Results

This section examines how effectively DPOS functions in real-world scenarios. The assessment evaluates performance in both normal and adverse conditions, as well as decision stability and overall cost efficiency. All experiments are designed to ensure comparability, statistical validity, and repeatability.

Evaluation Metrics

We use a combination of operational, economic, and stability metrics to assess decision quality in a manner that is understandable to both engineering and business stakeholders. The primary operational metric is the Authorization Success Rate (ASR), defined as the number of successful transactions divided by the total number of transaction attempts within a specified time period. ASR is reported for each channel (POS vs. online/mobile) and for each merchant group to illustrate how different policies impact various groups. The time from initiation (terminal action or gateway request) to completion (final authorization response) is called end-to-end latency. This duration is presented as distribution percentiles (P50, P95, and, under unstable conditions, P99). Percentile-based reporting reveals both typical behavior and tail latency, both of which are crucial for meeting SLAs. Cost per Transaction (CPT) measures economic efficiency by summing direct rail fees, authorization costs, settlement charges, SaaS overheads, and infrastructure costs spread over time. To enable comparison across rails and scenarios, all costs are converted to a single currency unit. All metrics include 95% confidence intervals calculated via bootstrap resampling.

Data Sources and Experimental Scenarios

The evaluation uses two publicly available, real-world datasets to characterize risk and latency. The Credit Card Fraud Detection dataset, which comprises 284,807 transactions over two days and 492 confirmed fraud

cases (fraud rate of 0.172%), serves as the source of transaction-risk data. Each record includes anonymized features, the transaction amount, the transaction time, and a binary fraud label. Fraud labels help determine the potential loss per transaction.

RIPE Atlas measurements, such as daily minimum RTT statistics and jitter distributions, along with publicly available RTT traces, are utilized to create network latency profiles. These figures help generate realistic latency distributions (P50/P95/P99) for different times of day and network conditions. There are four simulated payment methods: card payments, immediate transfers between accounts, online QR payments, and digital wallets that operate offline with limited risk and reconciliation upon reconnection. Stressful situations are used to evaluate DPOS under adverse and degraded conditions by deliberately augmenting the transaction stream with various failure and stress scenarios. These include DNS outages, modeled as 20–40-minute periods of high failure rates and RTT spikes; bandwidth throttling, resulting in high jitter and extended tail latency during peak mobile usage; issuer partial failures, which increase decline rates on the card network; and complete connectivity loss, which triggers controlled offline operation. All experiments are assessed against predefined SLA thresholds, including a minimum acceptable authorization success rate (ASR) and a maximum allowable P95 latency, ensuring objective and consistent measurement of SLA violations across different scenarios.

Evaluation Methodology

Evaluation follows a two-step process. First, offline replay maintains the original transaction order from the fraud dataset, with a latency profile derived from RIPE Atlas data applied to each transaction; to ensure fairness, competing policies are executed in the same sequence. Second, canary-style validation routes a small portion (10%) of synthetically reconstructed traffic through DPOS under feature flags to verify consistency between offline and quasi-online behavior. Statistical significance is assessed using Fisher's exact test or Z-tests for rate-based metrics such as ASR, and Mann-Whitney U tests for latency and cost distributions, with confidence intervals and p-values reported where appropriate. DPOS is compared to four basic strategies: Fixed Priority, where the order of the rails remains unchanged; Round-Robin, which chooses in a cycle without knowing about signals; Late Switch, which switches only after the SLA has been broken; and DPOS, as a full policy engine with risk budgeting, hysteresis, and cautious exploration. All baselines have the same limits, fees, and SLA thresholds.

Main Results

Based on an offline replay of the full sequence of 284,807 transactions from the Kaggle dataset (with a true fraud rate of 0.172%), the rail selector significantly outperformed the best baseline (the Late Switch policy) while also improving success, user experience, and cost efficiency. The authorization success rate (ASR) increased from 95.792% to 98.770%, an absolute gain of 2.978 percentage points and a roughly 71% relative reduction in transaction failures (from 4.208% to 1.230%). The 95th percentile latency decreased from 456.4 ms to 279.6 ms (about a 38.7% reduction), and the cost per transaction (CPT) dropped from 4.470 thousand rials to 2.794 thousand rials (approximately a 37.5% reduction). These improvements were achieved without any rise in SLA violations: the SLA breach rate per 10,000 transactions remained zero for both the best baseline and the rail selector. No risk budget overflow was observed, and due to hysteresis and switch penalties, the switching rate stayed effectively near zero, maintaining experience quality and operational stability.

At the level of network and acceptor incidents, the rail-selector's advantage remained consistent. During DNS disruptions (≈3% of traffic), ASR increased from 91.34% to 98.86%, a 7.52 percentage-point increase, while P95 latency decreased from 556.7 ms to 273.5 ms, representing approximately a 51% improvement. Under high throttle/jitter conditions (≈5% of traffic), ASR improved from 82.65% to 97.91%, a gain of 15.26 percentage points, and P95 latency decreased by roughly 47.5%. In issuer-partial outages, ASR increased from 94.94% to 98.83%, a 3.89 percentage-point increase, with a roughly 32% reduction in P95 latency. Even under normal conditions (≈90% of traffic), the rail-selector showed a 2.02 percentage-point improvement in ASR

and approximately a 28.5% reduction in P95 latency, demonstrating that its benefits are not limited to stress scenarios but also extend to daily network operation. Economically, the rail-selector reduced CPT across all scenarios by approximately 72% in DNS and throttle conditions and by around 37.5% in other cases, thanks to intelligent routing of stable traffic to lower-cost rails and avoidance of high-risk, low-yield paths.

Using ground-truth risk labels from the Kaggle dataset and real-world latency profiles from RIPE Atlas, the rail-selector not only enhances resilience during incidents (DNS, issuer, and throttle) and bridges performance gaps but also boosts economic efficiency and user experience under normal conditions, all without causing operational instability through excessive switching or exceeding risk budgets.

Table 6. Ablation Study (60k Subsample from the 284k Run)

ASR (%)	P95 (ms)	CPT (kIRR)	Switches / 1k	Configuration
98.818	279.6	2.796	0.0	Full (RailSelector)
98.822	278.9	2.796	0.0	− Risk Budget
98.760	279.2	2.796	0.0	− Switch Cost / Hysteresis

Table 6 presents an ablation study on a 60k subsample from the same 284k transaction run. The full configuration (rail selector) achieves an ASR of 98.818%, P95 latency of 279.6 ms, CPT of 2.796 kIRR, and zero switches per 1k transactions. Removing the risk-budget component yields nearly identical performance (ASR 98.822%, P95 278.9 ms), indicating that, under this specific threshold/budget setting, the available budget was adequate and no overflow occurred. Eliminating switch cost and hysteresis slightly reduces ASR (to 98.760%) without causing additional switching in this dataset; however, in more volatile environments, this removal is likely to significantly increase the switch rate per 1k transactions.

Table 7. Sensitivity to γ (Switch Cost) — 25k subset:

ASR (%)	P95 (ms)	CPT (kIRR)	Switches/1k	γ (Cost/Switch)
98.916	279.6	2.797	0.0	0
98.736	280.7	2.797	0.0	600
98.840	279.3	2.797	0.0	1200
98.844	281.4	2.797	0.0	2400
98.796	279.7	2.797	0.0	4800

Table 7 examines how sensitive the results are to the switching cost parameter γ using a 25k-sample. As γ varies from 0 to 4800 (cost units per switch), the ASR stays between 98.736% and 98.916%, the P95 latency remains close to 280 ms, CPT remains stable at about 2.797 kIRR, and the switch rate is nearly zero. In this case, a good operating point is around $\gamma \approx 1200$, which balances stability and efficiency. The minimal impact on switching shows the conservative default policy, which rarely switches rails; in more volatile datasets, the effect of γ would be more noticeable. With approximately 285k transactions, the observed gains in ASR (+2.98 percentage points) and P95 latency (−38.7%) are much larger than random noise, confirming the improvements are both statistically and practically significant. These findings suggest that DPOS improves resilience not only during infrastructure failures but also in normal operations. Gains are achieved without sacrificing stability or following the rules, demonstrating that combining multi-objective optimization, conservative learning, and rule-based safeguards is effective.

Conclusion and Future Work

This paper introduced DPOS, a dynamic and auditable policy engine designed for autonomous switching among various payment rails in high-throughput and disruption-sensitive environments. DPOS considers the

selection of payment rails as a multi-objective online optimization problem. This approach reduces transactional risk, economic costs, and end-to-end latency simultaneously while strictly adhering to operational and regulatory requirements. The framework achieves higher authorization success rates, significantly lower tail latency, and reduced costs per transaction in both normal and adverse conditions. It also maintains switching rates near zero and prevents SLA violations or risk budget breaches. These findings indicate that principled online decision-making, combined with conservative safeguards and auditability-by-design, can greatly enhance the reliability and efficiency of modern payment systems. From a business perspective, DPOS offers a practical deployment option. Its modular architecture allows for gradual feature rollouts using feature flags and canary traffic. Its deterministic fallback rules and limited offline modes enable graceful degradation during large-scale network failures. The audit framework is vital because it ensures every routing decision can be replicated and justified when reviewed from internal or external perspectives. Future research directions include: (1) augmenting the policy engine with formally verified constraints and safety proofs; (2) integrating causal models to more accurately measure the revenue impact of latency; (3) enhancing privacy protections through secure aggregation and differential telemetry; (4) incorporating safe reinforcement learning techniques to accelerate adaptation within strict risk boundaries; and (5) testing the framework in large-scale, multi-market deployments such as cross-border and real-time national payment rails.

References

Chou, K.-C., & Chen, P.-Y. (2024). *Dual-message QR codes*. Sensors, 24(10), 3055. https://doi.org/10.3390/s24103055

Güler, U., Atan, T., & Günneç, D. (2025). *Round-robin tournament scheduling under total game attractiveness objective*. In Proceedings of the 1st International Conference on Mathematics in Sport (Luxembourg, June 4–6).

Howell, B. (2025). *Beyond infrastructure: Internet ecosystem resilience and the public good*. Telecommunications Policy, 49(7). 10.1016/j.telpol.2025.102998

Hashimoto, K., & Konaka, E. (2025). *Optimization of the tournament format for the nationwide high school kyudo competition in Japan*. In Proceedings of the 1st International Conference on Mathematics in Sport (Luxembourg, June 4–6).

Morillo Reina, J. D., Mateo Sanguino, T. J. (2025). *Decentralized and secure blockchain solution for tamper-evident logging*. Future Internet, 17(3), 108. https://doi.org/10.3390/fi17030108

Nurmi, K., Kyngäs, J., & Järvelä, A. I. (2025). *Optimizing professional sports league games based on spectators and traveling*. In Proceedings of the 1st International Conference on Mathematics in Sport (Luxembourg, June 4–6).

Okubo, I., Hishiyama, R., Yamazaki, M., & Nakasako, S. (2024). *A study on the simple encryption of QR codes using random numbers and RC4 for fraud prevention in QR payments*. Electronics, 13(15), 3003. https://doi.org/10.3390/electronics13153003

Onumadu, P., Abroshan, H. (2024). *Near-field communication cyber threats and mitigation solutions in payment transactions: A review*. Sensors, 24(23), 7423. https://doi.org/10.3390/s24237423

Pacchiano, A., Ghavamzadeh, M., & Bartlett, P. L. (2024). *Contextual bandits with stage-wise constraints*. arXiv preprint arXiv:2401.08016. https://arxiv.org/abs/2401.08016

Slivkins, A., Zhou, X., Sankararaman, K. A., & Foster, D. J. (2024). *Contextual bandits with packing and covering constraints*. Journal of Machine Learning Research, 25, 1–56. https://www.jmlr.org/papers/v25/24-1220.html

Su, H., Niu, J., Liu, X., & Atiquzzaman, M. (2024). *SafeCoder: A machine-learning-based encoding system to embed safety identification information into QR codes*. Journal of Network and Computer Applications, 227, 103874. https://doi.org/10.1016/j.jnca.2024.103874

Yoon, J., & Kim, Y. (2025). *Offline payment of central bank digital currency based on a trusted platform module*. Journal of Cybersecurity and Privacy, 5(2), 14. https://doi.org/10.3390/jcp5020014

SEQUENTIAL TACTICAL OPTIMISATION IN INVASION SPORTS: A REGRET-BASED AND QUBO FORMULATION

Forbes, F.S.[a,c], Bracewell, P.J.[a,c]
[a] *Dovetale Sport*
[b] *University of Southern Queensland*
[c] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Decision-making in invasion sports such as rugby union, football, and basketball occurs as a sequence of tactical choices made under rapidly changing spatial and temporal conditions. During possession, teams repeatedly select between actions such as passing, carrying, or kicking while operating under defensive pressure and incomplete information. Although sports analytics has produced substantial advances in modelling match outcomes and evaluating individual events, relatively little research has examined how *sequences* of tactical decisions can be analysed and optimised within a unified mathematical framework.

This paper introduces a general framework for sequential tactical optimisation in invasion sports. Gameplay is modelled as a sequence of decision states defined by contextual variables such as field position, possession status, and match time. At each state, the team in possession selects from a discrete set of feasible tactical actions. An expected-value model is used to estimate the potential reward associated with each action, incorporating factors such as scoring probability, turnover risk, and territorial progression.

To evaluate decision quality independently of realised outcomes, the framework introduces regret-based metrics, defined as the difference between the expected value of the optimal action and the action actually chosen. This allows analysts to identify sub-optimal tactical decisions even when the observed outcome appears favourable, providing a principled basis for assessing decision-making under uncertainty.

To explore optimal decision sequences across multiple phases of play, the tactical decision problem is formulated as a Quadratic Unconstrained Binary Optimisation (QUBO) model. Binary variables correspond to candidate actions at each decision step, while penalty terms enforce feasibility constraints such as selecting exactly one action per state. The resulting formulation transforms tactical planning into a combinatorial optimisation problem solvable by classical or emerging quantum methods.

A proof-of-concept application in rugby union demonstrates how the framework reveals patterns of tactical inefficiency and identifies higher-value sequences of play. The approach provides a bridge between sports analytics and modern optimisation, offering a general foundation for sequential decision analysis across invasion sports.

Keywords: Sequential optimisation, regret metrics, QUBO, rugby union, decision-making under uncertainty

1.INTRODUCTION

1.1 Background

Invasion sports are characterised by continuous, high-frequency decision-making under uncertainty. At each moment of possession, teams must choose among competing tactical actions , passing, carrying, or kicking, while adapting to spatial constraints, defensive pressure, and evolving match context. These decisions occur within seconds and their cumulative quality directly bears on scoring opportunities and match outcomes.

Traditional performance analysis in rugby union and related sports relies heavily on retrospective video review and outcome-based metrics such as metres gained, possession time, and scoring events. While informative, such approaches conflate decision quality with outcome realisation and do not scale efficiently across large match datasets. A coach who selects the tactically optimal action may still concede possession through no fault of the decision itself; conversely, a poor tactical choice may yield a favourable result by chance.

Recent advances in sports analytics, including expected value (EV) models and machine learning, have improved predictive capabilities considerably (Odet et al., 2025; Watts & Pipping, 2025). However, these approaches largely evaluate discrete events in isolation rather than as components of a sequential decision process. The gap between event-level evaluation and multi-step optimisation remains largely unaddressed.

1.2 Problem Statement

The central limitation in current sports analytics is the absence of a unified framework that simultaneously: (i) evaluates decision quality independently of outcomes; (ii) accounts for sequential dependencies between decisions across successive phases of play; and (iii) enables optimisation of multi-step tactical sequences. Without such a framework, coaching insight remains partly subjective and post-hoc — tied to what happened rather than what should have happened.

1.3 Contributions

This paper makes four primary contributions:

- **A state-based formulation** of tactical decision-making in invasion sports, representing gameplay as a sequence of contextually defined decision states.
- **A regret-based evaluation framework** that measures the opportunity cost of observed decisions relative to the contextually optimal action, independent of realised outcomes.
- **A QUBO formulation** for sequential tactical optimisation, transforming multi-step decision planning into a combinatorial problem amenable to classical and quantum solvers.
- **A proof-of-concept application** in rugby union demonstrating how the framework identifies tactical inefficiency and generates actionable coaching insight.

Together, these contributions establish a unified framework linking evaluation, explanation, and optimisation of tactical decision-making in invasion sports.

2.RELATED WORK

2.1 Expected Value Models in Sports

Expected Points (EP) and win probability models assign a context-specific expected value to each tactical action, providing a structured basis for decision evaluation. Watts and Pipping (2025), in recent preprint work, introduce the first comprehensive EP framework for penalty decisions in Rugby Union, estimating expected value surfaces for kicking at goal versus kicking to touch. By comparing these surfaces they produce decision maps showing which choice maximises expected return under varying field position, score margin, and kicker accuracy. Their work demonstrates the tractability of EV modelling for specific rugby decisions but is limited to penalty scenarios and does not extend to broader in-play sequences or multi-step planning.

Parallel developments in basketball, notably Cervone et al. (2014), who model possession value as a continuous spatial stochastic process, demonstrate that EV frameworks generalise naturally across invasion sports. That work shows that multi-resolution state representations can meaningfully capture fast-moving play dynamics, and directly informs the zone-based state representation adopted in this paper.

2.2 Machine Learning in Rugby Analytics

Machine learning approaches have been widely applied to rugby performance prediction. Odet et al. (2025), using data from more than 2,000 matches, compare logistic regression, random forests, neural networks, and support vector machines for identifying winning strategies. Applying SHAP values for interpretability, they identify kicking performance, possession in attacking zones, ruck efficiency, and ELO-based team strength as the most influential outcome predictors. While producing actionable insights for coaches, the focus of that work is outcome prediction rather than decision evaluation; it identifies what matters but does not assess whether individual decisions were optimal or recommend alternative actions. Applied work more broadly has established that tactical choices such as kicking versus carrying can be represented quantitatively and linked to match outcomes (Watson et al., 2023; Mansurali & Mahmoud, 2024).

2.3 Sequential Decision Processes and Reinforcement Learning

Sequential decision problems in uncertain environments are most commonly formalised as Markov Decision Processes (MDPs), which define a framework for finding policies that maximise cumulative expected reward over time (Puterman, 1994). Reinforcement learning builds on this foundation, using iterative interaction with an environment to approximate optimal policies without requiring a fully specified transition model (Sutton & Barto, 2018). These approaches are powerful for long-horizon learning problems but are oriented toward policy optimisation through repeated simulation rather than explicit combinatorial optimisation of a finite observed sequence.

In the coaching context, where decisions must be evaluated post-hoc from match data rather than learned through repeated play, a direct combinatorial formulation is more tractable and interpretable. The QUBO approach adopted in this paper is complementary to MDP/RL methods: it optimises a specific observed decision sequence rather than learning a general policy, and produces solutions that are directly auditable by coaches. This positioning pre-empts the question of why combinatorial optimisation is preferred over policy learning for this specific application.

2.4 Regret in Sequential Decision Theory

Regret is a well-established concept in online learning and sequential decision theory. Mansour (2009) defines regret as the cumulative gap between the reward achieved by an algorithm and the reward of the best fixed action in hindsight. Gradu et al. (2023) extend this to adaptive regret for time-varying environments, which is particularly relevant to invasion sports where optimal actions change continuously with field position, momentum, and match state. Despite its theoretical maturity, regret has not previously been applied as an evaluation metric in tactical sports analytics. This paper introduces it as a principled, outcome-independent measure of decision quality.

2.5 QUBO Optimisation

Quadratic Unconstrained Binary Optimisation (QUBO) provides a versatile framework for combinatorial optimisation problems involving binary decision variables. Glover et al. (2019) provide a comprehensive tutorial showing how diverse constraint-based problems can be encoded as QUBO models via quadratic penalty terms, enabling solution by classical heuristics and quantum annealing hardware. The connection between QUBO and Ising spin-glass models, established by Lucas (2014), is particularly significant: it demonstrates that a wide class of NP-hard combinatorial problems, including sequencing problems directly analogous to tactical planning, can be encoded in the same Hamiltonian structure exploited by quantum annealers. Nakatsukasa et al. (2025) provide a recent practical demonstration of QUBO effectiveness in complex scheduling settings. To date, no published work has applied QUBO to tactical sports decision sequencing which is a gap this paper directly addresses.

2.6 Research Gap

Existing work addresses individual components,EV modelling, outcome prediction, regret theory, or combinatorial optimisation, but no framework integrates these into a single coherent analytical system for tactical sport. Furthermore, no prior work bridges the MDP/RL tradition with QUBO-based combinatorial optimisation for post-hoc coaching analysis of observed match sequences. This paper addresses both gaps.

3.MODELLING FRAMEWORK

3.1 Decision States

Gameplay is represented as a discrete sequence of decision states. At each moment t in which the team in possession must select a tactical action, the match context is characterised by a state vector:

$$s_t = (z_t,\ p_t,\ \tau_t,\ m_t)$$

where z_t denotes the field zone, p_t indicates possession state, τ_t captures match time, and m_t encodes momentum and contextual variables derived from recent match history. This representation balances rugby relevance with computational tractability while preserving the contextual factors most strongly associated with tactical norms.

3.2 Action Space

At each decision state the team in possession selects from a finite set of tactically meaningful actions:

$$a_t \in A(s_t)$$

The five core in-play actions modelled are: pass left, pass right, pick-and-go, quick tap, and kick to touch. These are defined consistently with the event taxonomy in the underlying tracking and event data, ensuring direct correspondence between the modelling framework and observed match behaviour.

3.3 Expected Value Function

The expected value of each action is estimated by a heuristic outcome model of the form:

$$V(s_t, a_t) = \alpha \cdot P_{score} - \beta \cdot P_{turnover} + \gamma \cdot \Delta WP$$

where P_score denotes the estimated scoring probability, P_turnover the turnover risk, and ΔWP the predicted change in win probability. Both P_score and P_turnover are modelled via sigmoid transformations of field zone and momentum indicators; ΔWP is derived as a weighted combination of these components. The coefficients α, β, and γ are calibrated heuristically in this proof-of-concept; data-driven calibration via match outcome regression is identified as a key direction for future work.

4.REGRET-BASED DECISION EVALUATION

4.1 Regret Definition

Regret at decision step t is defined as the difference between the expected value of the contextually optimal action and the expected value of the observed action:

$$R_t = \max_{a \in A(s_t)} V(s_t, a) - V(s_t, a_t^{obs})$$

This captures the opportunity cost of the observed tactical choice. By construction $R_t \geq 0$, with $R_t = 0$ indicating that the optimal action was selected. Higher values indicate greater sub-optimality.

4.2 Properties and Interpretation

The regret measure has several desirable properties for tactical analysis. It is outcome-independent: a decision yielding a poor result due to defensive excellence or chance will not be penalised if it represents the optimal expected-value choice in context. Conversely, a decision that yields a favourable result by chance, but was contextually sub-optimal, will register positive regret. This decoupling of decision quality from realised outcomes addresses a fundamental limitation of conventional outcome-based analysis.

Two operationalisations are employed. Threshold-based regret flags instances where the displacement of the action outcome falls below a minimum acceptable threshold (dx < 5 m in the current implementation). EV-based regret computes the continuous quantity R_t as defined above. Both are visualised in the decision dashboard to provide complementary diagnostic perspectives for coaching staff.

4.3 Cumulative Regret

Across a sequence of T decision steps, cumulative regret is defined as:

$$R_{total} = \sum_{t=1}^{T} R_t$$

Cumulative regret quantifies the aggregate opportunity cost of observed decision-making over an extended phase of play or a full match. Sequences or periods exhibiting high cumulative regret represent the highest-priority targets for coaching intervention.

5.QUBO FORMULATION FOR SEQUENTIAL OPTIMISATION

5.1 Binary Decision Variables

To extend optimisation across multiple phases of play, the decision problem is reformulated using binary variables. For each decision step t and each feasible action $a \in A(s_t)$, define:

$$x_{t,a} \in \{0, 1\}$$

where x_{t,a} = 1 if action a is selected at step t and x_{t,a} = 0 otherwise. The total number of binary variables is T × |A|, where T is the planning horizon and |A| is the action set cardinality.

5.2 Objective Function

The objective is to maximise total expected value across the decision sequence:

$$max \ \Sigma_t \ \Sigma_a \ V(s_t, a) \cdot x_{t,a}$$

Converting to the minimisation form required by QUBO solvers:

$$min \ -\Sigma_t \ \Sigma_a \ V(s_t, a) \cdot x_{t,a}$$

5.3 Feasibility Constraints

A feasibility constraint ensures that exactly one action is selected at each decision step:

$$\Sigma_a \ x_{t,a} = 1 \quad \forall t$$

This one-hot constraint is encoded as a quadratic penalty in the QUBO objective.

5.4 Full QUBO Formulation

Combining the objective and penalised constraint, the full QUBO model is:

$$min \ -\Sigma_{t,a} V(s_t, a) \cdot x_{t,a} + \lambda \cdot \Sigma_t (\Sigma_a x_{t,a} - 1)^2$$

where $\lambda > 0$ is the penalty parameter controlling the trade-off between objective maximisation and constraint satisfaction. In the current implementation $\lambda = 5.0$, selected to ensure constraint violations are sufficiently penalised relative to the EV objective. The resulting Q matrix is symmetric and sparse, with diagonal entries corresponding to individual action values and off-diagonal entries encoding constraint penalty interactions.

5.5 Solution and Interpretation

For moderate problem sizes ($T \leq 20$ steps, $|A| = 5$), the QUBO model is solved exactly using classical enumeration or greedy heuristics. For larger planning horizons, approximate solvers including simulated annealing and quantum annealing hardware provide scalable alternatives. The QUBO formulation is natively compatible with the D-Wave quantum annealing architecture, positioning this framework for direct quantum implementation as hardware accessibility improves.

The solution x* identifies the highest-value action sequence for a given planning horizon. Comparing x* against the observed sequence x^obs reveals the locations and magnitude of sub-optimal decisions, providing a coherent and actionable basis for coaching feedback. This complements policy-learning approaches such as reinforcement learning by enabling exact sequence optimisation from observed data, without requiring a simulation environment or repeated interaction.

6.DATA AND RUGBY CONTEXT

6.1 Data Sources

Two primary datasets underpin this research. Game tracking data provide frame-level positional and movement information for players and the ball, including spatial coordinates, velocity components, and direction of movement. Event data annotate discrete match events: passes, kicks, tackles, and scoring actions Each record is linked to a start frame, end frame, team in possession, and event type. Both datasets are recorded at a common frame rate, enabling direct temporal alignment.

6.2 Field Zone Representation

Field position is among the most important determinants of tactical behaviour in rugby union. The pitch is divided into five longitudinal zones reflecting coaching language and professional analysis conventions. The Defensive 22 (D22) prioritises safety and territory clearance; the Exit Zone rewards effective clearing while avoiding turnover; the Middle Zone supports a balance of possession and territorial progression; the Attacking

40 (A40) favours retention and structured attack; and the Red Zone, closest to the opponent's try line, strongly favours ball retention and phase construction. In the modelling framework, zone is treated as a categorical state variable and expected values are conditioned on zone to capture zone-specific risk-reward profiles.

6.3 Feature Engineering and Dataset Construction

Raw event records are aligned with tracking data and filtered to retain only moments of clear in-play tactical decision-making. Ruck events are treated as transitional phases between decision points rather than explicit decision instances. Key engineered features include field zone (categorical), normalised spatial coordinates consistent across teams and halves, a momentum indicator derived from recent event outcomes, and temporal context capturing match phase.

After filtering, cleaning, and feature construction, the final dataset consists of structured decision instances. Each instance records the contextual state, the observed action, and subsequent outcome information used to calibrate the EV model. This dataset underpins the regret analysis and QUBO optimisation in the proof-of-concept application.

7.PROOF-OF-CONCEPT APPLICATION IN RUGBY UNION

7.1 Experimental Setup

A proof-of-concept analysis is conducted using structured match data from professional rugby union. Decision instances are extracted and processed using the pipeline described in Section 6. The outcome model is applied to each instance to compute expected values for all five feasible actions given the contextual state. Regret is then calculated for each observed decision, and the QUBO model is applied to contiguous sequences of decision steps to identify optimal multi-step action plans.

7.2 Findings

Several consistent patterns emerge from the analysis. First, high-regret decisions cluster disproportionately in transition zones, particularly the Exit Zone and Attacking 40, where the range of EV differentials across actions is widest. These zones represent the highest-leverage contexts for coaching intervention. Second, a material proportion of outcome-successful plays exhibit positive regret: the observed action produced a favourable result, but a higher-EV alternative was available given the match context. This finding reinforces the core motivation of the regret framework and illustrates the inadequacy of outcome-only evaluation. Third, optimal sequences generated by the QUBO solver differ from observed sequences in ways that are interpretable and tactically coherent; for example, preferring phase-building in the Attacking 40 over early kicks to touch that sacrifice scoring probability for territory.

7.3 Dashboard Implementation

To support practical application, the analytical pipeline is integrated into an interactive coaching dashboard built using Dash and Plotly. The dashboard provides tabbed visualisation of event timelines, movement heatmaps, EV and regret metrics by decision instance, possession flow, and QUBO-derived action recommendations. A Decision Console tab presents coach-ready outputs including the optimal greedy action sequence, QUBO preview, and regret rankings, enabling direct use in pre- and post-match review.

8.DISCUSSION

8.1 Reframing Tactical Analysis

The proposed framework shifts the evaluative lens from "what happened?" to "what should have happened?" By measuring regret rather than outcome, the framework enables analysts to identify poor decisions even when they produce good results and to credit good decisions even when they produce poor results. This represents a meaningful advance in the precision and fairness of tactical evaluation, and aligns with broader trends in evidence-based coaching practice.

8.2 Practical Applications

The framework has direct practical value across several coaching contexts. In post-match review, regret rankings direct attention to the highest-priority decision errors, enabling more efficient and evidence-based video sessions. In player development, persistent high-regret patterns for specific players or game contexts identify

skill gaps amenable to targeted practice. In pre-match preparation, QUBO-generated optimal sequences provide a quantitative basis for game plan development under specific anticipated field positions and opponent tendencies.

8.3 Quantum Computing Relevance

The QUBO formulation is natively compatible with quantum annealing hardware such as the D-Wave system. As quantum computing matures and hardware becomes more accessible, the framework positions sports analytics to exploit quadratic speedups in combinatorial optimisation. For problems with long planning horizons or large action spaces, such as full-match tactical sequencing, quantum approaches may offer qualitative advantages over classical heuristics. The current implementation serves as a classical proof-of-concept that is directly portable to quantum hardware with no reformulation required.

8.4 Generalisability

Although the proof-of-concept is developed in rugby union, the framework is deliberately general. Any invasion sport with a discrete action space and context-dependent tactics — including association football, basketball, and rugby league — can be modelled within the same state-action-value structure. The key requirements are aligned tracking and event data with sufficient resolution to construct decision instances, and domain knowledge to define the action taxonomy and zone structure.

9. LIMITATIONS AND FUTURE WORK

Several limitations of the current implementation are acknowledged. The outcome model is heuristic and relies on manually calibrated parameters; future work will replace it with a model trained directly on match outcome data, improving both predictive validity and zone-action sensitivity. The action space (five actions) does not capture the full richness of in-play decisions such as line breaks, offloads, or structured set-piece variants. The planning horizon in the current QUBO implementation is constrained by classical solver tractability; quantum annealing would substantially relax this limitation. Finally, empirical validation using a larger multi-season dataset with expert-labelled decisions would provide stronger grounds for external validation of the regret metric.

A paired extension of this work targeting a full quantum annealing implementation is in preparation. That paper will evaluate computational performance of quantum hardware on tactical decision problems of varying horizon length and action space cardinality, directly building on the QUBO formulation established here.

10. CONCLUSION

This paper introduces a unified framework for evaluating and optimising sequential tactical decision-making in invasion sports. By integrating an expected value outcome model, regret-based decision metrics, and a QUBO optimisation formulation, the framework provides a novel lens through which tactical performance can be understood; one that is sensitive to decision quality independently of realised outcomes.

Applied to rugby union, the proof-of-concept demonstrates that high-regret decisions cluster in tactically critical zones, that outcome-successful plays can still exhibit decision inefficiency, and that QUBO-derived optimal sequences provide coherent and actionable coaching recommendations. The framework is general, extensible to other invasion sports, and positioned for direct integration with quantum optimisation hardware. This work establishes a foundation for a new class of decision-support tools in elite sport: tools that evaluate not just what happened, but whether the right choice was made.

References

Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. Journal of the American Statistical Association, 111(514), 585–599. doi:10.1080/01621459.2016.1141685

Glover, F., Kochenberger, G., & Du, Y. (2019). Quantum Bridge Analytics I: A tutorial on formulating and using QUBO models. 4OR, 17(4), 335–371. doi:10.1007/s10288-019-00424-y

Gradu, P., Hazan, E., & Minasyan, E. (2023). Adaptive regret for control of time-varying dynamics. Proceedings of Machine Learning Research, 211, 560–572.

Lucas, A. (2014). Ising formulations of many NP problems. Frontiers in Physics, 2, Article 5. doi:10.3389/fphy.2014.00005

Mănescu, D. C. (2025). Big data analytics framework for decision-making in sports performance optimization. Data, 10(7), 116. doi:10.3390/data10070116

Mansour, Y. (2009). Regret minimization: Algorithms and applications [Lecture slides]. New York University.
Mansurali, A., & Mahmoud, A. B. (2024). Sports analytics: Data-driven sports and decision intelligence. In Sports Analytics: Data-Driven Sports and Decision Intelligence (pp. 1–17). Springer. doi:10.1007/978-3-031-63573-1_1
Nakatsukasa, I., Nakano, K., Parque, V., & Ito, Y. (2025). Optimizing heat treatment schedules via QUBO formulation. Applied Sciences, 15(16), 8847. doi:10.3390/app15168847
Odet, A., Bechard, T., Moretto, P., Dejean, S., & Pasquaretta, C. (2025). A machine learning and explainability-driven methodology for identifying winning strategies in rugby union. Decision Analytics Journal, 15, 100568. doi:10.1016/j.dajour.2025.100568
Puterman, M. L. (1994). Markov decision processes: Discrete stochastic dynamic programming. Wiley.
Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
Watson, N., Hendricks, S., Stewart, T., & Durbach, I. (2021). Integrating machine learning and decision support in tactical decision-making in rugby union. Journal of the Operational Research Society, 72(10), 2274–2285. doi:10.1080/01605682.2020.1779624
Watts, K., & Pipping-Gamón, J. (2025). Kicking for goal or touch? An expected points framework for penalty decisions in rugby union. arXiv:2512.00312.

A QUBO FORMULATION FOR LATENT POSSESSION SEQUENCE RECONSTRUCTION IN RUGBY UNION

Forbes, F.S.[a,b] Bracewell, P.J.[a,c]
[a] *Dovetale Sport*
[b] *University of Southern Queensland*
[c] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Ball possession in Rugby Union is commonly treated as a frame-wise classification problem derived from manual annotation, ball-tracking systems, or local predictions from player-tracking features. However, possession is more naturally understood as a temporally persistent and spatially structured latent state, with transitions occurring under tactical, territorial, and phase-of-play constraints. Frame-independent classifiers may therefore produce implausible rapid possession switches during contested breakdowns, even when they contain useful local evidence.

This paper proposes a methodological extension to prior work on possession inference from player tracking data by formulating possession reconstruction as a globally constrained binary optimisation problem. Building on evidence that collective movement features such as Centre of Play, Movement of Play, spatial dispersion, and smoothed trajectory descriptors contain possession-related signals, we define possession as a latent binary sequence over time rather than as a set of independent frame-level labels.

The proposed framework constructs an energy function incorporating four components: local classification evidence, territorial alignment derived from Centre of Play, collective spatial structure derived from Movement of Play, and temporal transition penalties that discourage implausible short-duration possession reversals. This energy function can be expressed as a Quadratic Unconstrained Binary Optimisation (QUBO) problem, making it compatible with classical simulated annealing and, subject to embedding and hardware constraints, quantum annealing solvers.

The contribution of this paper is conceptual and methodological. It does not claim empirical superiority over existing models, but instead sets out a reproducible framework for converting frame-level possession probabilities into globally coherent possession sequences, providing a bridge between rugby analytics, structured latent-state inference, and combinatorial optimisation.

Keywords: Binary quadratic optimisation; Temporal state estimation; Latent state inference; Spatio-temporal modelling; Quantum annealing

1. INTRODUCTION

Ball possession is a foundational analytical construct in Rugby Union, underpinning event detection, phase counting, territory analysis, and strategic review. Despite its importance, possession is difficult to observe directly from player-tracking data. Breakdowns, rucks, mauls, and contested phases create conditions in which the ball is obscured, contested, or stationary, making frame-level classification an inherently noisy proposition. Even high-performing classifiers can produce implausible rapid possession reversals during these phases, undermining downstream analyses that depend on stable, coherent possession timelines.

Prior work by Wei et al. (2026) demonstrated that collective movement features extracted from player-tracking data carry reliable possession-related signals. Features such as Centre of Play, Movement of Play, spatial dispersion, and smoothed trajectory descriptors were shown to distinguish possession states at the frame level with meaningful accuracy using a Random Forest classifier. However, frame-level classifiers do not, by construction, enforce temporal consistency across adjacent observations. A sequence of independent classification decisions can yield possession sequences that oscillate implausibly on timescales shorter than any realistic tactical transition.

This limitation motivates a fundamentally different formulation of the possession inference problem. Rather than treating each frame in isolation, this paper proposes that possession be recovered as a globally coherent latent binary sequence, where the full trajectory is optimised simultaneously subject to evidence from local classifiers, territorial context, and a penalty on implausible transitions. This is a structured prediction problem, and its natural representation is as a Quadratic Unconstrained Binary Optimisation (QUBO) problem.

The QUBO framework is compatible with classical solvers such as simulated annealing and, subject to embedding and hardware constraints, with quantum annealing platforms such as D-Wave systems. Although empirical validation using quantum hardware is beyond the scope of this paper, the formulation is specified in a form that admits submission to such solvers.

The contribution of this paper is not a new classifier, nor an empirical claim of quantum advantage. It is a formal QUBO construction that converts noisy frame-level possession evidence into a temporally coherent latent

possession sequence. Specifically, the paper: (i) frames rugby possession inference as a global sequence reconstruction problem rather than a classification task; (ii) proposes a QUBO energy function incorporating local classification evidence, territorial alignment, collective spatial structure, and temporal smoothness; and (iii) sets out a reproducible evaluation blueprint for comparing this approach against frame-level and HMM baselines.

2. BACKGROUND AND RELATED WORK

2.1 Possession Inference from Player Tracking Data

Automated possession inference from player-tracking data has attracted growing attention as high-resolution positional data becomes more widely available in rugby match analysis. Tracking systems provide frame-level coordinates for all players on the field, enabling the extraction of collective spatial features that would be inaccessible from event-log data alone.

Wei et al. (2026) demonstrated that a range of movement and spatial features derived from tracking data carry reliable possession signals. Centre of Play, defined as the spatial centroid of the tracked players, shifts systematically with possession, as teams in possession tend to draw players forward and toward the ball. Movement of Play captures the direction and magnitude of collective displacement, which differs between attacking and defending configurations. Spatial dispersion and compression metrics reflect formation dynamics that differ between possession and non-possession states. A Random Forest classifier (Breiman, 2001) trained on these features achieved meaningful discrimination between possession states, establishing that the signal exists and can be extracted without direct ball tracking.

This paper inherits Wei's feature framework as the source of local classification evidence and proposes a method for converting those local predictions into a globally coherent sequence.

2.2 Temporal Smoothing and Latent-State Models

The temporal structure of sporting sequences has been widely modelled using Hidden Markov Models (HMMs). An HMM assumes that observations are generated by an underlying latent state sequence governed by Markov transition dynamics. In the possession context, the latent states correspond to possession by the home team or the opposition, and observations correspond to frame-level feature vectors or classifier outputs. HMMs are appealing because they enforce temporal consistency through the Markov assumption, and efficient inference is available via the Viterbi algorithm (Rabiner, 1989).

Conditional Random Fields (CRFs) extend this idea to discriminative structured prediction, allowing the model to learn feature weights that jointly maximise the probability of the entire sequence given the observations (Lafferty et al., 2001). CRFs have been applied successfully in activity recognition and event segmentation in sport, where temporal dependencies between adjacent labels are informative.

A shared limitation of HMMs and CRFs is that they typically operate within a probabilistic graphical model framework, which assumes parametric emission and transition distributions. The approach proposed here is complementary: rather than fitting a generative or discriminative model, we define an energy function over binary sequences and seek the minimum-energy configuration, which is a combinatorial optimisation problem. CRFs represent a natural comparator but are not included in the initial evaluation blueprint of Section 5; the present formulation is designed to test whether explicit energy minimisation over a hand-specified, interpretable QUBO objective provides value before introducing learned sequence-level feature weights.

2.3 QUBO and Global Optimisation

A Quadratic Unconstrained Binary Optimisation problem is defined over a vector of binary variables $z = (z_1, z_2, ..., z_T) \in \{0,1\}^T$, with objective:

$$\text{minimise} \quad z^T Q z$$

where Q is a real-valued upper-triangular matrix encoding both linear terms (diagonal entries) and pairwise interaction terms (off-diagonal entries). The diagonal entry Q_{ii} represents the cost of setting $z_i = 1$, and the off-diagonal entry Q_{ij} represents the cost of the joint assignment $(z_i=1, z_j=1)$.

QUBO problems are NP-hard in general but admit a wide range of approximate solvers. Classical approaches include simulated annealing (Kirkpatrick et al., 1983), tabu search, and branch-and-bound methods. A comprehensive treatment of QUBO modelling and solver strategies is provided by Glover et al. (2019). Quantum annealing, as implemented on D-Wave hardware, solves QUBO problems natively by encoding the binary variables as qubit spins and the Q matrix as an Ising Hamiltonian, then evolving the system toward its ground state (Lucas, 2014; Kadowaki & Nishimori, 1998).

The binary structure of the possession inference problem, where $z_t \in \{0, 1\}$ for each time frame t, maps naturally to the QUBO format. This is the key motivation for the formulation proposed here.

3. PROBLEM FRAMING

Let a match be represented as a sequence of T frames indexed by t = 1, ..., T, where each frame corresponds to a fixed temporal interval (e.g., one second) of player-tracking data. We define the latent possession sequence as a binary vector:

$$z = (z_1, z_2, ..., z_T) \in \{0, 1\}^T$$

where $z_t = 1$ indicates that the focal team is in possession at frame t, and $z_t = 0$ indicates that the opposition holds possession. Possession is treated as exhaustive and mutually exclusive at each frame, consistent with the operational definition used in annotated match data.

The aim is not to maximise frame-level classification accuracy in isolation, but to recover a sequence z that is simultaneously consistent with local classification evidence, coherent with territorial context, and free of implausible rapid possession reversals. This is a sequence reconstruction problem rather than a pointwise prediction problem.

Let $p_t \in (0, 1)$ denote the frame-level probability that the focal team is in possession at frame t, as output by a classifier trained on collective movement features following Wei et al. (2026). The QUBO objective combines these local probabilities with pairwise and contextual terms to define a global energy function E(z) that is minimised over all possible binary sequences.

Formally, we seek:

$$z^* = \underset{z \in \{0,1\}^T}{\operatorname{argmin}} E(z)$$

where E(z) is defined in the following section.

4. PROPOSED QUBO FORMULATION

4.1 Local Evidence Term

The local evidence term encodes the frame-level possession probabilities from the Random Forest classifier. For each frame t, the unary cost of assigning $z_t = 1$ is the negative log-probability of possession, and the cost of assigning $z_t = 0$ is the negative log-probability of non-possession. The unary contribution to the energy function is:

$$E_1(z) = \sum_t [\, -\log(p_t) \cdot z_t \; - \; \log(1 - p_t) \cdot (1 - z_t) \,]$$

which simplifies, up to a constant, to:

$$E_1(z) = \sum_t \; (\log(1 - p_t) - \log(p_t)) \cdot z_t$$

The coefficient of z_t is negative when $p_t > 0.5$, encouraging $z_t = 1$, and positive when $p_t < 0.5$, encouraging $z_t = 0$. These unary terms populate the diagonal of the Q matrix: $Q_{tt} = \log(1 - p_t) - \log(p_t)$.

4.2 Temporal Smoothness Term

Possession in Rugby Union does not switch at the frame rate of the tracking system. Transitions occur at discrete tactical events: kicks, turnovers, lineouts, scrums, and penalties. A possession sequence that switches between frames is inconsistent with the physiology and tactics of the game. To penalise implausible rapid switches, we introduce a pairwise transition penalty:

$$\psi(z_t, z_{t+1}) = \lambda \cdot I(z_t \neq z_{t+1})$$

where $I(\cdot)$ is the indicator function and $\lambda > 0$ is a hyperparameter controlling the cost of each transition. This penalty can be expressed in QUBO form by noting that $I(z_t \neq z_{t+1}) = z_t + z_{t+1} - 2z_t z_{t+1}$, giving the pairwise contribution:

$$E_2(z) = \lambda \cdot \sum_t (z_t + z_{t+1} - 2\, z_t\, z_{t+1})$$

The linear terms λz_t and λz_{t+1} modify the diagonal entries of Q, and the interaction term $-2\lambda z_t\, z_{t+1}$ populates the off-diagonal entry $Q_{t,t+1} = -2\lambda$. This term encourages runs of consistent possession labels and discourages single-frame or short-duration reversals.

4.3 Territorial Alignment Term

Field position is a reliable contextual cue for possession. Teams in possession drive the Centre of Play (CoP) forward; teams without possession typically cede territory. Let $c_t \in [-1, 1]$ denote the direction-normalised field position of the CoP at frame t, where positive values indicate position toward the focal team's attacking end and negative values indicate position toward their defensive end. All territorial and movement variables are assumed to be direction-normalised so that positive values consistently

represent the focal team's attacking direction. We define a territorial evidence term:

$$E_3(z) = -\alpha \cdot \sum_t c_t z_t$$

where $\alpha > 0$ is a weighting hyperparameter. When $c_t > 0$ the CoP is in the attacking half, making the coefficient of z_t negative and encouraging $z_t = 1$. When $c_t < 0$ the CoP is in the defensive half, encouraging $z_t = 0$. When $c_t = 0$ the term is neutral. This term is added to the diagonal of Q: $Q_{tt} \mathrel{+}= -\alpha\, c_t$.

4.4 Collective Structure Term

Movement of Play (MoP) captures the direction and magnitude of collective displacement. During some attacking phases, collective movement is expected to become more directionally aligned toward the opposition's try line, while defensive or contested phases may exhibit different dispersion and movement signatures. Let $m_t \in [-1, 1]$ denote the signed MoP for the focal team at frame t, where positive values indicate forward collective movement. The optional collective structure term is:

$$E_4(z) = \beta \cdot \sum_t (-m_t) \cdot z_t$$

where $\beta \geq 0$ is a weighting hyperparameter. When $m_t > 0$ (focal team moving forward), the coefficient of z_t is negative, encouraging $z_t = 1$. This term is added to the diagonal of Q: $Q_{tt} \mathrel{+}= -\beta\, m_t$.

4.5 QUBO Matrix Construction

The full energy function is the sum of the four component terms:

$$E(z) = E_1(z) + E_2(z) + E_3(z) + E_4(z)$$

This can be written compactly as:

$$E(z) = z^T Q\, z$$

where the Q matrix is assembled as follows. For interior frames $t = 2, \ldots, T-1$, each frame has two adjacent neighbours, so the linear contribution from the smoothness term appears twice. The diagonal entry is:

$$Q_{tt} = \log(1-p_t) - \log(p_t) + 2\lambda - \alpha\, c_t - \beta\, m_t$$

For the first and final frames, each has only one adjacent neighbour, so the smoothness contribution is λ rather than 2λ:

$$Q_{11} = \log(1-p_1) - \log(p_1) + \lambda - \alpha\, c_1 - \beta\, m_1$$

$$Q_{TT} = \log(1-p_T) - \log(p_T) + \lambda - \alpha\, c_T - \beta\, m_T$$

For each adjacent pair (t, t+1), the off-diagonal entry is:

$$Q_{t,t+1} = -2\lambda$$

All other entries of Q are zero. The matrix Q is real-valued, upper-triangular, and of dimension $T \times T$. The global optimisation problem minimise $z^T Q z$ over $z \in \{0,1\}^T$ is the QUBO that is submitted to the solver.

5. PROPOSED EVALUATION DESIGN

5.1 Data Requirements

A complete empirical evaluation would require a dataset of rugby union matches with synchronised player-tracking coordinates and manual possession annotations. Specifically, the following components are required: (i) frame-level two-dimensional player coordinates for all tracked players at sufficient temporal resolution (for example, 10 Hz or higher); (ii) manual possession labels for each frame or possession phase, produced by trained analysts following a documented annotation protocol; (iii) restart annotations identifying scrums, lineouts, penalties, and kick restarts, which define the ground-truth transition events against which transition detection performance can be assessed; and (iv) match metadata including team identities, match period, and score differential, which may serve as additional contextual features.

The annotation protocol should distinguish clearly between clear-possession phases, contested phases (rucks and mauls where possession is genuinely ambiguous), and dead-ball periods, to enable appropriate handling of ground-truth uncertainty.

5.2 Baseline Models

Three models should be compared in future empirical work. The Random Forest frame-level classifier following Wei et al. (2026) provides the baseline unary evidence and also serves as an independent comparison model. An HMM with two latent states (possession and non-possession) provides a principled temporal baseline that enforces Markov consistency and can be trained using the Baum-Welch algorithm and decoded using Viterbi. The QUBO global optimisation model proposed here provides the experimental condition. Comparisons between these three baselines isolate the contribution of temporal structure (HMM vs. RF) and the contribution of global optimisation (QUBO vs. HMM).

5.3 Evaluation Metrics

Performance should be assessed across five dimensions. Frame-level discrimination should be measured by area under the ROC curve (AUC) or balanced accuracy, to account for class imbalance during contested phases. Transition volatility should be measured as the number of possession changes per unit time in the inferred sequence, compared to the annotated ground truth; a well-behaved sequence should produce transitions at a rate consistent with the tactical structure of the match. Phase-level consistency should measure the proportion of annotated possession phases that are recovered as contiguous runs in the inferred sequence. Agreement with restart annotations should measure whether inferred possession transitions co-occur with annotated restart events, which represent the ground-truth trigger for possession change. Finally, interpretability and coaching usability should be assessed qualitatively, evaluating whether the inferred possession timeline is legible to rugby analysts and suitable for use in tactical review.

5.4 Hyperparameter Sensitivity

The three hyperparameters λ, α, and β should be tuned by cross-validation across match segments. The smoothness parameter λ represents a fundamental trade-off: low values allow the sequence to follow frame-level evidence closely, at the risk of producing volatile sequences; high values enforce temporal stability, at the risk of over-smoothing and missing genuine transitions. The territorial weight α should be tuned by evaluating whether CoP provides additional information beyond the classifier output. The spatial structure weight β is optional and may be set to zero if MoP does not independently improve sequence quality. Sensitivity analysis should report performance metrics across a range of λ values to characterise this trade-off empirically.

6. ILLUSTRATIVE EXAMPLE

To ground the formulation, we trace through a small synthetic sequence of T = 20 frames. The frame-level possession probabilities below are representative of a passage of play in which the focal team holds clear possession, passes through a contested ruck phase, then regains possession:

$$p = (0.91, 0.89, 0.85, 0.82, 0.54, 0.47, 0.52, 0.49, 0.83, 0.88,$$
$$0.90, 0.86, 0.88, 0.44, 0.41, 0.46, 0.43, 0.87, 0.91, 0.93)$$

Frames 5–8 and 14–17 represent contested phases where the classifier output oscillates near 0.5. Under independent thresholding at $p_t = 0.5$, the inferred possession sequence is:

$$z^{RF} = (1,1,1,1,1,0,1,0,1,1,1,1,1,0,0,0,0,1,1,1)$$

This sequence contains four single-frame or two-frame possession reversals (frames 6, 8, 14–17), none of which would correspond to a legitimate rugby transition event such as a turnover or restart. The QUBO formulation addresses this by penalising each transition. Setting $\lambda = 0.5$ and $\alpha = \beta = 0$ (classifier and smoothness terms only, no territorial or MoP weighting), the diagonal entries of Q are:

$$Q_{tt} = \log(1-p_t) - \log(p_t) + 2\lambda \quad \text{(interior frames)}$$

For frame 6, $p_6 = 0.47$, so independent thresholding assigns $z_6 = 0$, creating a single-frame reversal within an otherwise possession-dominant passage. The local evidence for switching is weak: $\log(0.53) - \log(0.47) \approx 0.12$. With $\lambda = 0.5$, switching at frame 6 incurs two transition penalties, one entering and one leaving the reversal, with total cost $2\lambda = 1.0$. The smoothness penalty therefore outweighs the local gain from switching, and the QUBO solution retains $z_6 = 1$. The QUBO-smoothed sequence recovers:

$$z^{QUBO} = (1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1,1,1)$$

The four-frame contested window at frames 14–17 is retained as a transition because the evidence is consistently below 0.5 across four consecutive frames, making the transition cost affordable relative to the accumulated local evidence. By contrast, the two isolated reversals at frames 6 and 8 are suppressed because their energy cost (the negative log-probability gain from flipping) does not outweigh the smoothness penalty λ.

This toy example illustrates the key property established analytically in Section 7: short-duration oscillations in contested phases are suppressed by the smoothness term without eliminating genuine multi-frame transitions. The precise threshold depends on λ and the depth of the probability dip, and would be calibrated empirically on annotated match data in future work.

7. ANALYTICAL PROPERTIES OF THE FRAMEWORK

This section describes properties of the QUBO objective function that follow directly from its mathematical structure. These are formal consequences of the energy definition, not empirical predictions.

When $\lambda = 0$, the temporal smoothness term vanishes and the objective decomposes into T independent scalar minimisations. The QUBO solution is then equivalent to thresholding the frame-level classifier probabilities at $p_t = 0.5$, recovering exactly the independent classifier. In the limit $\lambda \to \infty$, the transition cost dominates all local

evidence and the optimiser collapses to a constant sequence ($z_t = c$ for all t, where $c \in \{0,1\}$). Intermediate values of λ interpolate between these two extremes, trading local discriminability against temporal coherence. These are properties of the objective function, not claims about empirical performance.

The territorial alignment term $-\alpha\, c_t\, z_t$ acts as a contextual prior. When $c_t > 0$, the term reduces Q_{tt}, making $z_t = 1$ less costly; when $c_t < 0$, it increases Q_{tt}, making $z = 1$ more costly. The parameter α controls how strongly field position is weighted relative to classifier evidence. Setting $\alpha = 0$ disables territorial weighting, making the framework a two-component (classifier + smoothness) model.

Similarly, the collective structure term $-\beta\, m_t\, z_t$ acts as a directional prior. Setting $\beta = 0$ removes the MoP contribution entirely, and the framework degrades gracefully to the simpler formulation. The modular structure of the energy function means that each component can be included or excluded independently, enabling systematic ablation in future empirical work.

Together, these properties mean the framework has well-defined limiting behaviour: it reduces to independent classification when all contextual and smoothness weights are zero, and it admits a direct comparison with frame-level and HMM baselines through systematic variation of λ, α, and β.

8. DISCUSSION

8.1 Why Global Optimisation Matters

Possession inference is fundamentally a sequence reconstruction problem, not a collection of independent classification problems. The tactical structure of Rugby Union imposes strong temporal constraints on possession sequences: possessions persist for multiple phases; transitions are triggered by discrete events; and contested periods create structured ambiguity that local classifiers cannot resolve without reference to their temporal context. A classifier that is evaluated frame-by-frame is solving a different problem from an analyst who needs a coherent possession timeline for tactical review.

The QUBO framework reframes possession inference in terms that align with this analytical need. By jointly optimising the full sequence, it respects the global consistency requirements that frame-level classifiers ignore. The energy function is explicit about the trade-offs being made: how much weight to give local evidence, how much to penalise transitions, and how much to trust territorial context. This explicitness is itself analytically valuable, as it makes the assumptions of the model transparent and testable.

8.2 Relationship to Wei's Work

This paper is an explicit extension of Wei et al. (2026). Wei et al. established that possession signal exists in collective movement features and that a Random Forest classifier can extract it with meaningful accuracy at the frame level. That contribution is the prerequisite for the present work: the QUBO framework inherits the frame-level probabilities from Wei et al.'s classifier as its primary input. Wei et al. establish that possession signal exists; this paper proposes how to impose temporal coherence on that signal.

The two contributions are therefore complementary rather than competing. A natural next step is to implement both as components of a unified pipeline: Wei et al.'s classifier generates unary evidence, and the QUBO solver converts that evidence into a globally coherent sequence. Empirical comparison of the two models on the same annotated dataset would directly quantify the benefit of global optimisation over independent classification.

8.3 Quantum Annealing as a Future Pathway

The QUBO formulation proposed here is compatible with quantum annealing hardware, including D-Wave systems. Quantum annealing solves QUBO problems by encoding binary variables as qubit spins and exploiting quantum tunnelling to explore the solution landscape more efficiently than classical thermal annealing in certain regimes (Kadowaki & Nishimori, 1998; Johnson et al., 2011). The Q matrix defined in Section 4.5 could be submitted to a D-Wave solver after appropriate scaling, embedding, and solver-parameter selection.

We make no claim that quantum annealing has been applied in this work, nor that it would outperform classical simulated annealing for the problem sizes arising in match analysis. However, the formulation is mathematically specified and compatible with QUBO solvers, including classical simulated annealing and quantum annealing hardware, and as quantum annealing systems continue to scale in qubit count and connectivity, larger and more complex energy functions may become tractable. The sporting analytics domain provides a practical testbed for evaluating the utility of near-term quantum hardware on structured optimisation problems.

8.4 Practical Relevance

Coaching and performance analysis in Rugby Union depend on stable, coherent possession timelines. Event detection algorithms that identify phase boundaries, linebreak opportunities, and territory transitions require possession sequences as inputs. Momentum analysis, which tracks the ebb and flow of territorial and tactical

control, is ill-defined without reliable possession labels. Comparative analysis across matches or seasons, which requires consistent labelling of possession phases, is undermined by noisy frame-level classifications.

The QUBO framework addresses these practical needs by specifying a mechanism for producing possession sequences that are globally consistent and free of the rapid fluctuations that arise from independent frame-level classification. The tunable hyperparameters allow practitioners to customise the trade-off between sensitivity and stability, depending on the analytical application. A high-smoothness setting may be appropriate for season-level statistical analysis, while a lower smoothness setting may be preferred for fine-grained event detection where precision near transition boundaries is critical.

9. LIMITATIONS AND FUTURE WORK

Several limitations bound the scope of this contribution. No empirical implementation of the QUBO pipeline is reported; the energy function has been specified mathematically and the evaluation framework has been defined, but neither has been applied to annotated match data. No quantum hardware results are claimed. The properties described in Section 7 are analytical consequences of the objective function and require empirical testing to determine whether they translate into practical performance improvements on real rugby data.

Future work should pursue four priorities. First, the pipeline should be implemented end-to-end: Wei's classifier should be used to generate unary evidence from player-tracking data, the Q matrix should be constructed as described in Section 4.5, and a classical simulated annealing solver should be used to find approximate solutions. Second, the implemented pipeline should be evaluated on an annotated match dataset using the metrics defined in Section 5.3. Third, the RF, HMM, and QUBO models should be compared empirically to quantify the contribution of global optimisation. Fourth, the computational feasibility of submitting the QUBO to D-Wave hardware should be assessed, including embedding efficiency and solution quality relative to classical baselines.

10.CONCLUSION

This paper has proposed a QUBO formulation for latent possession sequence reconstruction in Rugby Union. Building on prior work demonstrating that collective movement features carry possession-related signal at the frame level, the framework converts frame-level classification probabilities into a globally coherent binary possession sequence by minimising an energy function that combines local evidence, territorial alignment, collective spatial structure, and temporal transition penalties within a single objective.

The formulation is mathematically specified and compatible with QUBO solvers, including classical simulated annealing and quantum annealing hardware. The Q matrix is fully defined, the hyperparameters are interpretable with clear limiting behaviour, and the evaluation blueprint provides a reproducible pathway for future empirical validation. The contribution is a formal methodological foundation for globally optimised possession inference: one that respects the temporal structure of Rugby Union, connects to the analytical needs of coaches and performance staff, and identifies a concrete pathway for evaluating quantum annealing on structured inference problems in sports analytics.

Acknowledgements

The authors acknowledge the contributions of Qunwei Wei, whose foundational work on possession inference from player-tracking data motivated this extension.

References

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Glover, F., Kochenberger, G., & Du, Y. (2019). Quantum Bridge Analytics I: A tutorial on formulating and using QUBO models. 4OR, 17(4), 335–371. https://doi.org/10.1007/s10288-019-00424-y

Johnson, M. W., Amin, M. H. S., Gildert, S., Lanting, T., Hamze, F., Dickson, N., Harris, R., Berkley, A. J., Johansson, J., Bunyk, P., Chapple, E. M., Enderud, C., Hilton, J. P., Karimi, K., Ladizinsky, E., Ladizinsky, N., Oh, T., Perminov, I., Rich, C., Thom, M. C., Tolkacheva, E., Truncik, C. J. S., Uchaikin, S., Wang, J., Wilson, B., & Rose, G. (2011). Quantum annealing with manufactured spins. Nature, 473, 194–198. https://doi.org/10.1038/nature10012

Kadowaki, T., & Nishimori, H. (1998). Quantum annealing in the transverse Ising model. Physical Review E, 58(5), 5355–5363. https://doi.org/10.1103/PhysRevE.58.5355

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. Science, 220(4598), 671–680. https://doi.org/10.1126/science.220.4598.671

Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the 18th International Conference on Machine Learning, 282–289.

Lucas, A. (2014). Ising formulations of many NP problems. Frontiers in Physics, 2, Article 5. https://doi.org/10.3389/fphy.2014.00005

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257–286. https://doi.org/10.1109/5.18626

Wei, Q., Bracewell, P. J., McGahan, B., & Joseph, J. (2026). Possession without the ball: Latent-state inference from collective movement in rugby union. Proceedings of the 18th ANZIAM MathSport Conference.

REFORMULATING LOGISTIC CLASSIFICATION MODELS AS QUADRATIC OPTIMISATION PROBLEMS: A QUANTUM ANNEALING FRAMEWORK FOR TACTICAL DECISION MODELLING IN RUGBY

Emmy Gauper [a,b], Paul J. Bracewell [a,d], F. Sam Forbes [a,c]
[a] *Dovetale Sport*
[b] *Victoria University of Wellington*
[c] *University of Southern Queensland*
[d] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Recent advances in quantum annealing have expanded the range of optimisation problems that can be expressed and solved in quadratic binary form. In sport analytics, however, most predictive models remain framed as probabilistic classification tasks rather than energy minimisation problems. This paper establishes a structural bridge between these two paradigms by demonstrating how logistic regression models with binary predictors can be reformulated as Quadratic Unconstrained Binary Optimisation (QUBO) problems suitable for quantum annealing.

Using ruck turnover prediction in rugby union as a motivating case study, we show that the log-odds function of a logistic model can be directly mapped to a QUBO energy function when predictors are appropriately discretised. Linear coefficients correspond to QUBO linear terms, and interaction effects naturally generate quadratic couplings. We further extend the framework to Constrained Quadratic Models (CQMs), enabling integer and continuous decision variables with feasibility constraints via hybrid classical-quantum solvers.

Due to limited structured ruck data, open-source soccer phases-of-play data were used to prototype the optimisation pipeline. Exact classical solvers were employed to validate structural feasibility in the absence of quantum hardware access. Results confirm that turnover probability minimisation can be equivalently expressed as energy minimisation. However, unconstrained optimisation may yield tactically infeasible solutions, highlighting the necessity of embedding domain-specific constraints. This work does not claim computational advantage over classical methods. Rather, it provides a formal mathematical pathway for transforming predictive sport models into quadratic optimisation problems, establishing a foundation for future real-time tactical optimisation using hybrid and quantum solvers.

Keywords: Logistic regression, QUBO, Quantum annealing, Hybrid optimisation, Tactical decision modelling

1. INTRODUCTION

Predictive modelling in sport analytics has matured considerably over the past decade, with logistic regression remaining a workhorse method for binary outcome prediction: win or loss, turnover or retention, tackle made or missed. These models are well understood, statistically interpretable, and straightforward to fit. Yet their conventional role ends at prediction. Once trained, a logistic model reports the probability of an outcome given observed conditions; it does not, in itself, identify which combination of controllable inputs minimises the probability of an adverse outcome.

Separately, quantum annealing has emerged as a promising computational paradigm for combinatorial optimisation. Unlike classical gradient-based methods that traverse a solution landscape sequentially, quantum annealers exploit quantum tunnelling to search the solution space in parallel, making them particularly well-suited to problems with large discrete search spaces (D-Wave, 2025b). The canonical input format for quantum annealers is the Quadratic Unconstrained Binary Optimisation (QUBO) problem, in which an objective function composed of linear and quadratic binary terms is minimised over all binary assignments.

These two paradigms have not previously been connected formally in the sport analytics literature. To our knowledge, this is the first explicit algebraic mapping from logistic regression models to QUBO formulations in a sports analytics context. This paper establishes that connection and derives its implications. We demonstrate that a logistic regression model fitted to binary or discretised predictors can be algebraically reformulated as a QUBO energy function, with log-odds coefficients mapping directly to QUBO linear terms and interaction effects generating quadratic couplings. The reformulation is mathematically exact: minimising the QUBO energy function is equivalent to finding the combination of binary predictor values that minimises the predicted

log-odds of the adverse outcome. This transforms the logistic model from a passive prediction instrument into an active decision instrument.

We ground this framework in a practically important problem: ruck turnover prediction in rugby union. The ruck is the most frequent event in a match, accounting for approximately 42% of all match events (O'Muilleoir, 2024), yet remains comparatively under-studied analytically. The ruck presents a natural binary decision structure: each candidate player can either join or not join a forming ruck, yielding a combinatorial search space that grows as 2^n with squad size and motivates optimisation methods beyond brute force enumeration.

Due to limited rugby-specific data at the time of prototyping, open-source soccer phase-of-play data from the 2024/2025 Australian A-League season (SkillCorner, 2025) is used to demonstrate the full pipeline. The soccer dataset is used solely to validate the structural correctness of the mapping and optimisation pipeline; no domain-level conclusions are drawn from it. All domain-specific conclusions are drawn from rugby literature and the rugby pilot dataset.

The remainder of this paper is structured as follows. Section 2 reviews background on logistic regression, QUBO formulations, and quantum annealing in sport. Section 3 derives the formal logistic-to-QUBO mapping. Section 4 extends the framework to CQMs for mixed-variable problems. Section 5 presents prototype results. Section 6 discusses implications for ruck modelling and the necessity of feasibility constraints. Section 7 concludes.

2. BACKGROUND

2.1 Logistic Regression in Sport Analytics

Logistic regression models the probability that a binary outcome $Y = 1$ occurs given a predictor vector $X = (X_1, \ldots, X_p)$. The multiple logistic regression model takes the form:

$$log[p(X) / (1 - p(X))] = \beta 0 + \beta 1X1 + \cdots + \beta pXp$$

where $p(X) = Pr(Y = 1 \mid X)$, $\beta 0$ is the intercept, and $\beta 1, \ldots, \beta p$ are regression coefficients. The left-hand side is the log-odds or logit. Since $p(X)$ is monotone increasing in the log-odds, minimising the log-odds is equivalent to minimising the predicted probability of $Y = 1$ (James et al., 2013). This equivalence is the cornerstone of the QUBO mapping derived in Section 3.

In rugby union, Hendricks et al. (2018) identified the principal predictors of ruck outcome: player counts at the ruck, advantage-line progression, ball-carrier falling direction, and active ball placement. Kraak et al. (2017) established that retention rates exceed 90% in elite competition, confirming the rarity of turnovers and the class imbalance challenge that any ruck prediction model must address.

2.2 Quadratic Unconstrained Binary Optimisation

A QUBO problem seeks to minimise an objective function of the form:

$$min\ E(x) = \Sigma i\ ai\ xi + \Sigma i\ \Sigma j{>}i\ bij\ xi\ xj + c$$

where $x = (x_1, \ldots, x_n) \in \{0, 1\}^n$ are binary decision variables, a_i are linear (diagonal) coefficients, b_{ij} are quadratic (off-diagonal) coupling coefficients, and c is a constant offset. The problem is NP-hard in general, but quantum annealers solve it by mapping the QUBO matrix Q onto a physical Ising spin Hamiltonian and applying quantum tunnelling to search the energy landscape (D-Wave, 2022). The Ising and QUBO formulations are mathematically equivalent under the substitution $s_i = 2x_i - 1 \in \{-1, +1\}$.

The QUBO matrix Q is defined with $Q_{ii} = a_i$ along the diagonal and $Q_{ij} = b_{ij}$ in the upper triangle. The objective can then be written compactly as $E(x) = x^T Q\ x + c$, a form that maps directly onto the dimod BinaryQuadraticModel API (D-Wave, 2025c).

2.3 Quantum Annealing in Sport

Applications of quantum annealing in sport analytics remain limited and are dominated by scheduling and player selection problems. Pechler and Phillipson (2025) applied quantum annealing to multi-league sports scheduling, implementing four formulations but finding that quantum approaches did not yet outperform classical benchmarks in computation speed. Fujii and Matsui (2021) demonstrated that a QUBO-based solver outperformed Mixed-Integer Quadratic Programming (MIQP) for round-robin tournament scheduling.

In player selection, Iturrospe (2021) optimised soccer team line-up decisions using a quantum annealer, while Gündüz and Yilmaz (2022) applied similar techniques to fantasy basketball selection. Toda et al. (2024) extended player grouping optimisation to multiplayer games. Feld et al. (2020) applied quantum annealing to archetypal analysis of athletes. Critically, no published work has applied quantum annealing to rugby union, and no published work has connected logistic regression outputs to QUBO inputs for real-time tactical decision support in any sport. The present paper addresses both gaps.

3. FROM LOGISTIC REGRESSION TO QUBO: THE FORMAL MAPPING

3.1 Binarisation of Predictors

The QUBO formulation requires all decision variables to be binary. When the logistic regression model contains continuous or integer predictors, these must first be discretised into binary indicator variables. For a continuous predictor X with domain [a, b], the domain is partitioned into K intervals and K indicator variables $x_1, \ldots, x_K$ defined, where x_k = 1 if X ∈ [c_{k−1}, c_k) and 0 otherwise. Exactly one indicator variable takes the value 1 in any observation. The original coefficient β is replaced by K position-specific coefficients $\beta_1, \ldots, \beta_K$ estimated from the data.

In the prototype, x_start (ball x−coordinate at phase start, range [−60, 60]) was discretised into four 30−unit segments (x_start_1 through x_start_4), x_end similarly, and minute_start into five 10−minute segments (minute_10 through minute_50). This yields 19 binary predictors from three original variables. The choice of interval width introduces a bias–variance trade−off: finer discretisation preserves more representational fidelity but expands the binary predictor count and the QUBO matrix dimension, while coarser discretisation reduces the search space at the cost of within−interval precision.

3.2 Proposition 1: The Logistic–QUBO Equivalence

Proposition 1 (Logistic–QUBO Equivalence). Let $\beta 0, \beta 1, \ldots, \beta p$ be the fitted coefficients of a logistic regression model with binary predictors $x \in \{0, 1\}^p$. Define the QUBO matrix Q with $Q_{ii} = \beta i$ along the diagonal and $Q_{ij} = \beta ij$ (pairwise interaction coefficients, if present) in the upper triangle, and constant offset $c = \beta 0$. Then minimising the QUBO energy $E(x) = x^T Q\, x + c$ over $x \in \{0, 1\}^p$ is equivalent to minimising the predicted probability $p(x) = \sigma(L(x))$ of the outcome Y = 1 over the same binary domain.

Proof sketch. Since $\sigma(\cdot)$ is strictly monotone increasing, minimising p(x) is equivalent to minimising L(x). Since:

$$L(x) = \beta 0 + \beta 1 x1 + \ldots + \beta pxp = \beta 0 + \Sigma i\, \beta i\, xi = x^T Q\, x + c = E(x)$$

the two objectives are identical.

The mapping from logistic model to QUBO is therefore:

- Intercept $\beta 0 \rightarrow$ QUBO constant offset $c = \beta 0$.
- Linear coefficient $\beta i \rightarrow$ QUBO diagonal term $ai = \beta i$ for all $i = 1, \ldots, p$.
- Off-diagonal terms $bij = 0$ in the absence of interaction terms.

When the logistic model includes pairwise interaction terms βij xi xj, these extend the log-odds to:

$$L(x) = \beta 0 + \Sigma i\, \beta i\, xi + \Sigma i\, \Sigma j{>}i\, \beta ij\, xi\, xj$$

which maps exactly onto the full QUBO objective with bij = βij. The QUBO matrix Q is the upper-triangular matrix of all logistic regression coefficients (linear on the diagonal, pairwise interactions above it), and minimising $x^T Q\, x + c$ is exactly equivalent to finding the binary predictor configuration that minimises the predicted turnover probability. This is the central result of the paper.

3.3 Implementation via dimod

The mapping is implemented using D-Wave's dimod library (D-Wave, 2025c). The BinaryQuadraticModel class accepts linear coefficients as a dictionary and quadratic coefficients as a dictionary of variable pairs. The fitted logistic regression coefficients are passed as linear terms, and the intercept is set as an offset:

```
bqm = dimod.BinaryQuadraticModel('BINARY')
for var, coef in coeffs.items():
```

```
        bqm.add_variable(var, coef)
    bqm.offset = intercept
```

The resulting BQM is solver-agnostic: it can be submitted to dimod's ExactSolver (brute force, for validation), SimulatedAnnealingSampler (classical heuristic), or D-Wave's quantum hardware via the LeapHybridSampler or DWaveSampler. Only the sampler specification requires modification to transition from classical validation to quantum hardware deployment.

3.4 Scope and Limitations

The QUBO mapping is exact under three conditions: (a) all predictors are binary, (b) the logistic model contains at most pairwise interaction terms, and (c) the objective is minimisation of log-odds. Higher-order interactions cannot be accommodated, as QUBO is limited to quadratic terms. In practice, three-way interactions are rarely included in sport analytics models.

A further limitation is that the QUBO minimum identifies the binary configuration with the lowest predicted log-odds, which may be tactically infeasible. This is addressed in Section 4 (CQM constraints) and Section 6 (feasibility discussion). Finally, the quality of the QUBO solution depends entirely on the quality of the underlying logistic model; model misspecification propagates directly into the optimisation.

4. EXTENSION TO CONSTRAINED QUADRATIC MODELS

4.1 Motivation and Structure

The pure QUBO formulation requires all decision variables to be binary. In the ruck context, several variables of practical importance are naturally integer or continuous: match timing (minute), spatial coordinates (Xruck, Yruck), and player counts (ownTeamNum, oppTeamNum). Forcing these into binary indicator sets through discretisation introduces information loss and can expand the variable count substantially.

D-Wave's Constrained Quadratic Model (CQM) framework (D-Wave, 2025a) extends QUBO by permitting binary, integer, and continuous variables within a single objective, subject to equality and inequality constraints. The CQM minimises:

$$\Sigma i\ ai\ xi\ +\ \Sigma i \leqslant j\ bij\ xi\ xj\ +\ c$$

subject to constraints h_k(x) = 0 or g_k(x) ≤ 0, where x may include variables of any type. All variables must have finite bounds. The CQM is solved by D-Wave's LeapHybridCQMSampler, which combines classical and quantum computation, allocating subproblems to quantum hardware where computationally advantageous.

4.2 CQM Formulation for Ruck Optimisation

For the ruck application, the CQM supports the following variable structure simultaneously:

- Binary variables xi ∈ {0, 1} representing the join decision for each candidate player position i ∈ {hooker, tight-head prop, lock 1, lock 2, blindside flanker, number 8}.
- Integer variables for match timing (minute_start ∈ [0, 80]) and ruck spatial coordinates (Xruck ∈ [0, 100], Yruck ∈ [0, 70]).
- A cardinality constraint Σi xi ⩽ n_max, bounding the total number of players committing to the ruck and preserving defensive coverage elsewhere.
- A mutual exclusivity constraint Σk phase_k = 1 over the set of binary phase-type indicators, ensuring exactly one phase type is active.

The logistic regression coefficients for each variable type are passed as the objective coefficients in the CQM, and the constraint structure is specified separately. This clean separation between objective (predictive model) and constraints (tactical knowledge) is a key architectural feature of the CQM approach: it allows the logistic model to be updated independently of the feasibility constraints as data accumulate.

4.3 Interaction Terms as Quadratic Couplings

Pairwise interactions between player join decisions can be incorporated as quadratic coupling terms bij xi xj in the objective, encoding the synergistic or antagonistic effects of co-presence at the ruck. For example, if the joint presence of the hooker and flanker is particularly effective at securing possession, their interaction coefficient

b_{hooker, flanker} would be negative (reducing the log-odds of turnover beyond the sum of their individual effects). Conversely, if both locks joining simultaneously is redundant, their coupling term may be positive.

Estimating reliable interaction coefficients requires sufficient cross-tabulated data. With the current pilot dataset of seven rucks, interaction estimation is not feasible. However, the CQM objective function accommodates these terms directly, and the R implementation of the logistic model can include interaction terms via standard syntax (e.g., ownTeamNum * oppTeamNum) once data support their estimation.

5. PROTOTYPE IMPLEMENTATION AND RESULTS

5.1 Data and Logistic Model Specification

The prototype was implemented using phases-of-play data from five matches of the 2024/2025 Australian A-League soccer season (SkillCorner, 2025). The outcome variable was team_possession_loss_in_phase (Y = 1: possession lost). An initial logistic regression model was fitted in R using all available predictors, then reduced by removing variables with $p > 0.05$. The retained continuous-predictor model was:

$$\begin{aligned}\text{log-odds} = {} & -60.36 - 0.011 \times \text{minute_start} + 3.249 \times \text{phase_chaotic} \\ & + 0.908 \times \text{phase_create} + 1.922 \times \text{phase_direct} + 1.128 \times \text{phase_finish} \\ & + 1.404 \times \text{phase_quick_break} + 2.084 \times \text{phase_set_play} \\ & + 0.674 \times \text{phase_transition} - 0.027 \times \text{x_start} + 0.041 \times \text{x_end}\end{aligned}$$

The large negative intercept (−60.36) reflects the rarity of possession loss. Phase type was the dominant predictor: chaotic phases ($\beta = 3.249$, $p < 0.001$) and set plays ($\beta = 2.084$, $p < 0.001$) were most strongly associated with loss. Match minute was negatively associated with loss (possession more stable later in matches). The x_end coefficient (+0.041) indicates that phases ending in retreating positions increased turnover probability.

For the QUBO prototype, continuous predictors were binarised. Table 1 shows the resulting binary model coefficients.

Variable	Coefficient (β)	Description
(Intercept)	−0.876	Baseline (build-up phase, minutes 41–80, far end)
minute_10	+0.501**	Play in first 10 minutes
minute_20	+0.575***	Play in minutes 11–20
minute_40	+0.496**	Play in minutes 31–40
x_start_1	+1.021**	Phase starts in deep defensive third
x_start_2	+0.685*	Phase starts in middle third
x_end_1	−2.904***	Phase ends in deep defensive third (protective)
x_end_2	−2.331***	Phase ends in middle third
x_end_3	−1.930***	Phase ends in attacking third
phase_chaotic	+3.355***	Chaotic phase type (highest loss risk)
phase_direct	+2.689***	Direct play phase
phase_set_play	+1.575***	Set play phase
phase_transition	+1.179*	Transition phase

Table 1: Binary logistic regression coefficients for QUBO formulation (selected predictors). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Reference levels: build-up phase, minutes 41–80.

5.2 QUBO Prototype

The QUBO matrix was constructed by placing the binary logistic regression coefficients along the diagonal, with all off-diagonal entries set to zero (no interaction terms included in this prototype). The intercept was included as a constant offset. The dimod ExactSolver enumerated all $2^{17} = 131{,}072$ binary configurations to identify the global minimum.

The minimum energy configuration was:

```
minute_30 = 1,  phase_transition = 1,  x_start_3 = 1,  x_end_1 = 1
All other variables = 0
```
Energy (objective): −2.537 Predicted probability of loss: p ≈ 0.073

The optimal QUBO configuration identifies a transition phase commencing in minutes 21–30, starting in the attacking third and ending in the deep defensive third. The predicted possession loss probability of approximately 7% is the minimum achievable under this model. However, this configuration is not operationally feasible: a transition phase starting in the attacking third and ending in the defensive third is impossible by definition. This infeasibility is a direct consequence of omitting domain constraints from the QUBO formulation, demonstrating that the QUBO mapping correctly identifies the mathematical minimum of the log-odds surface while simultaneously illustrating that this minimum may lie in a region of the binary state space that is unreachable in practice.

5.3 Hybrid CQM Prototype

The CQM prototype used integer variables for x_start and x_end, bounded to [−15, 15] due to the computational cost of exact classical solving with the full range [−60, 60]. The dimod ExactCQMSolver was used, with a sum−to−one constraint on phase type indicators. The minimum was:

```
minute_start = 51,  x_start = 15,  x_end = −15,  phase_transition = 1
```
Objective (z): −61.270 Predicted probability of loss: p ≈ 2.46×10^{-27}

The effectively zero predicted loss probability reflects the dominance of the large negative intercept combined with the negative coefficient on minute_start. The recommended strategy (transition phase starting near the opponent's goal at minute 51 and ending at the opposite end) is operationally impossible. This confirms that unconstrained optimisation over a logistic model surface produces solutions that are mathematically correct but tactically vacuous without feasibility constraints. Importantly, both prototypes validate the structural correctness of the logistic-to-quadratic pipeline.

6. IMPLICATIONS FOR RUGBY RUCK MODELLING

6.1 The Ruck Join Decision as a Native QUBO Problem

The ruck join decision is structurally a binary optimisation problem. For each of n candidate players available when a ruck forms, define $x_i \in \{0, 1\}$ as the indicator that player i joins. Once a rugby−specific logistic regression model is estimated for ruck turnover probability with player join indicators as predictors, the QUBO mapping of Section 3 directly provides the optimal join configuration. With n = 6 tracked positions (as in the pilot dataset), the search space is $2^6 = 64$—trivially enumerable. At full squad scale with n = 15, the $2^{15} = 32{,}768$ configuration space, combined with real−time constraints, begins to justify faster optimisation methods.

6.2 Embedding Feasibility Constraints

The prototype results demonstrate that feasibility constraints are not optional—they are essential to actionable recommendations. For the ruck application, three categories of constraint are required.

Physical constraints from the laws of the game (World Rugby, 2025) require that any player joining the ruck does so from behind the offside line, and that players beyond a threshold distance from the ruck at formation cannot realistically arrive in time. Proximity constraints of the form $x_i = 0$ for $d_i > d_{max}$ can be enforced as QUBO penalties (setting a_i to a large positive value M) or as CQM equality constraints.

Tactical constraints reflect team strategy and field coverage requirements. The total number of players committing to the ruck is bounded: too few may concede the ruck, too many leaves the defensive structure exposed. A CQM constraint $n_{min} \leq \Sigma_i x_i \leq n_{max}$ enforces both bounds simultaneously. Position−specific constraints can encode the rule that the scrumhalf typically does not join (instead preparing to receive ball), reducing the effective binary search space.

Comparative analysis constraints unlock an alternative use of the framework. Rather than seeking the globally optimal join configuration, the QUBO objective can be evaluated for any two feasible configurations to rank them by predicted turnover probability. This supports in-session coaching decisions of the form: given that the ruck is at field position X and player i is committed, does adding player j lower or raise turnover probability?

6.3 Preliminary Empirical Evidence from the Rugby Pilot Dataset

The rugby pilot dataset, drawn from seven sequential rucks in a match between St Patrick's College and Wellington College, provides preliminary evidence directionally consistent with the modelling framework. In all observed instances where the defending team outnumbered the attacking team by two or more players, the attacking team failed to retain possession. Where the attacking team achieved numerical parity or superiority, possession was retained in over 80% of instances. While the small sample precludes statistical inference, these patterns align closely with Hendricks et al. (2018), who found that three to five active defensive players reduced retention probability by 85% in international competition. Player imbalance (ownTeamNum − oppTeamNum) is therefore the priority predictor variable for inclusion in the QUBO objective once sufficient data are available.

6.4 Transition to Production

The Python code developed for this project is directly portable to D-Wave quantum hardware: the ExactSolver and ExactCQMSolver are drop-in replacements for DWaveSampler and LeapHybridCQMSampler respectively, requiring only the sampler instantiation line to change. The dimod library provides a consistent API across all solver backends, minimising the engineering barrier to quantum hardware deployment once D-Wave API access is established.

The full production pathway requires three developments in sequence: (1) assembly of a multi-match, multi-team rugby ruck dataset enabling reliable logistic model estimation; (2) formal specification and encoding of tactical feasibility constraints informed by coaching staff; and (3) integration with real-time player tracking to provide live xi availability indicators as each ruck forms.

7. CONCLUSION

This paper has derived a formal mathematical bridge between logistic regression classification models and Quadratic Unconstrained Binary Optimisation. The central result—stated as Proposition 1—is that when a logistic regression model is fitted to binary predictors, its log-odds function is structurally identical to a QUBO energy function: linear coefficients map to QUBO diagonal terms, pairwise interaction coefficients map to off-diagonal couplings, and the intercept maps to the constant offset. Minimising the QUBO energy is therefore mathematically equivalent to identifying the binary predictor configuration that minimises the predicted probability of the adverse outcome.

The framework has been implemented using D-Wave's dimod library and validated on open-source soccer data as a structural proxy for the rugby ruck prediction problem. The extension to CQMs accommodates mixed-variable logistic models and critically enables the embedding of domain-specific feasibility constraints, addressing the finding that unconstrained optimisation over a logistic model surface can produce recommendations that are tactically impossible.

Applied to ruck turnover prediction in rugby union, the framework provides a rigorous and implementable pathway from a fitted statistical model to a quantum-annealing-compatible decision optimiser. The binary structure of player join decisions maps directly onto QUBO binary variables, and the exponential growth of the search space with squad size provides the fundamental computational motivation for quantum and hybrid solvers as the framework scales.

This work explicitly does not claim current computational advantage of quantum over classical methods for this problem. Rather, it establishes the formal mathematical and software foundation required to investigate that question when quantum hardware access becomes available. The generalisation is immediate: any sport analytics problem expressible as a logistic regression over binary or discretised predictors is, by the mapping derived here, expressible as a QUBO problem amenable to quantum annealing—establishing a general pathway for converting probabilistic classification models into optimisation frameworks over controllable inputs.

Acknowledgements
The first author acknowledges the supervision of Dr Paul Bracewell and Dr Ivy Liu at Te Herenga Waka - Victoria University of Wellington. The authors thank SkillCorner for making their phases-of-play data publicly available, and D-Wave Quantum Inc. for open-source access to the dimod and Ocean SDK libraries.

References

D-Wave Quantum Inc. (2022). Problem formulation guide. https://www.dwavequantum.com/media/bu0lh5ee/problem-formulation-guide-2022-01-10.pdf

D-Wave Quantum Inc. (2025a). Leap service's hybrid solvers. https://docs.dwavequantum.com/en/latest/industrial_optimization/leap_hybrid.html

D-Wave Quantum Inc. (2025b). What is quantum annealing? https://docs.dwavequantum.com/en/latest/quantum_research/quantum_annealing_intro.html

D-Wave Quantum Inc. (2025c). dimod: A shared API for samplers. https://docs.dwavequantum.com/en/latest/ocean/api_ref_dimod/index.html

Feld, S., Roch, C., Geirhos, K., & Gabor, T. (2020). Approximating archetypal analysis using quantum annealing. In Proceedings of the 28th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2020) (pp. 292–296).

Fujii, K., & Matsui, T. (2023). Solving break minimization problems in mirrored double round-robin tournament with QUBO solver. arXiv:2307.00263.

Gündüz, S., & Yılmaz, İ. (2024). Harnessing the power of quantum computing to build an ideal team and line-up for the Euroleague Fantasy Challenge. International Journal of Innovative Approaches in Science Research, 8(2), 56–69. https://doi.org/10.29329/ijiasr.2024.1054.2

Hendricks, S., van Niekerk, T., Sin, D. W., Lambert, M., den Hollander, S., Brown, J., Maree, W., Treu, P., Till, K., & Jones, B. (2018). Technical determinants of tackle and ruck performance in international rugby union. Journal of Sports Sciences, 36(5), 522–528.

Iturrospe, A. (2021). Optimizing decision making for soccer line-up by a quantum annealer. arXiv:2112.13668.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

Kraak, W., Coetzee, F., & Venter, R. (2017). Analysis of the general match profile of international rugby union between 2007 and 2013. International Journal of Performance Analysis in Sport, 17(3), 303–318.

Ó Muilleoir, C. (2024). Advancing player modelling in rugby union: An evaluation of path analysis techniques [Dissertation/project report]. National College of Ireland.

Pechler, O., & Phillipson, F. (2025). Toward quantum annealing for multi-league sports scheduling. In Innovations for Community Services: 25th International Conference, I4CS 2025, Munich, Germany, June 11–13, 2025, Proceedings (pp. 159–188). Springer.

SkillCorner. (2025). Phases of play open data. GitHub. https://github.com/SkillCorner/opendata

Toda, K., Inoue, S., & Tobe, Y. (2024). Optimization of player-combinations in multiplayer games. In 2024 International Conference on Information Networking (ICOIN) (pp. 746–750). IEEE.

World Rugby. (2025). Law 15: Ruck. https://passport.world.rugby/laws-of-the-game/laws-by-number/15-ruck/

EVALUATING TACTICAL DECISION QUALITY IN RUGBY UNION USING REGRET-BASED METRICS

Maw, Khin Myo Thant [a,b], Bracewell, P.J. [a,d], Forbes, F.S. [a,c]
[a] *Dovetale Sport*
[b] *Victoria University of Wellington*
[c] *University of Southern Queensland*
[d] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Rugby union is characterised by rapid tactical decision-making under uncertainty. During open play, players must repeatedly choose between alternative actions such as passing, carrying, or kicking while operating under defensive pressure and incomplete information. Traditional performance analysis evaluates these actions primarily through their outcomes, such as metres gained, turnovers, or scoring events. However, outcome-based statistics do not necessarily reflect decision quality, as favourable outcomes may arise from sub-optimal decisions, and optimal decisions may fail due to defensive pressure or stochastic events.

This paper introduces a decision-evaluation framework for rugby union based on regret metrics, a concept widely used in sequential decision theory but rarely applied in sport analytics. The framework models each in-play moment as a decision instance defined by contextual state variables such as field position and game state, together with a discrete set of feasible tactical actions. An expected-value model estimates the outcome potential associated with each action. Regret is then defined as the difference between the expected value of the optimal action and the action actually chosen. A case study using integrated event and tracking data demonstrates how regret metrics reveal hidden patterns of decision inefficiency across field zones and match contexts.

Keywords: Sequential decision theory, Expected value modelling, Game-state modelling, Regret metrics

1. INTRODUCTION

Rugby Union is characterised by continuous decision-making under uncertainty. During a match, players must quickly decide whether to pass, kick, or carry the ball, and these decisions affect territory, scoring chances, and the risk of losing possession. Many choices are made within seconds and under strong defensive pressure.

Traditional coaching decisions are reviewed through video analysis conducted after the match. While useful, this approach is inherently subjective and does not scale well across many games. It also focuses primarily on outcomes, such as tries or turnovers, rather than evaluating whether the decision itself was sound at the moment it was made. Sport analytics has developed rapidly in sports such as football, basketball, and baseball, incorporating tracking data, event data, and machine learning models. By contrast, rugby analytics has made slower progress, particularly in the area of decision quality and computational optimisation. This paper addresses that gap by introducing regret analysis and Quadratic Unconstrained Binary Optimisation (QUBO) into rugby decision analytics. The central argument is that decision quality and match outcome are distinct quantities that must be measured separately. A poor outcome does not necessarily imply a poor decision, and a sound decision may fail due to randomness or superior opposition defence. Without a formal method for evaluating decision quality, coaches are forced to rely on subjective judgement and outcome-based statistics that may systematically misattribute success and failure.

The remainder of this paper is structured as follows. Section 2 reviews related work. Section 3 describes the rugby context and field representation. Section 4 outlines the data sources and preparation. Section 5 presents the methodology, including the outcome model, regret metrics, and QUBO formulation. Section 6 reports results from a case study application. Section 7 discusses findings and Section 8 concludes.

2. RELATED WORK

2.1 Machine Learning and Explainability in Rugby

Machine learning has been widely applied in rugby union to analyse performance and predict match outcomes. Odet et al. (2025) propose a machine-learning and explainability-driven framework for identifying winning strategies in professional rugby, benchmarking logistic regression, random forests, neural networks, and support vector machines across more than 2,000 matches. Their results identify kicking performance, possession in

attacking zones, ruck efficiency, and ELO-based team strength as the most influential factors for match outcomes, with SHAP values providing interpretable explanations at global, team, and match levels.

However, the focus of this work is prediction rather than decision evaluation. While it explains which factors are important, it does not assess whether individual tactical decisions were optimal, nor does it recommend alternative actions. These limitations motivate the use of regret metrics and optimisation in the present research.

2.2 Expected Points and Tactical Decision Modelling

Expected Points (EP) models provide a structured way to evaluate rugby decisions by assigning an expected value to alternative actions. Watts and Pipping (2025) introduce the first comprehensive EP framework for penalty decisions in rugby union, estimating decision surfaces for kicking at goal versus kicking to touch. By comparing these surfaces, they construct decision maps showing which option maximises expected return under varying field positions, score differences, and kicker accuracy.

Despite its strengths, that framework is limited to penalty decisions and does not extend to broader in-play tactical choices or multi-step decision problems. The present research addresses these gaps by extending expected value concepts to general in-play actions and combining them with optimisation.

2.3 Regret Metrics in Sequential Decision-Making

Regret metrics provide a formal measure of decision quality. Mansour (2009) defines regret as the difference between the reward of the chosen action and the best possible action in hindsight. This concept allows decision quality to be evaluated independently of random outcomes. More recent work by Gradu et al. (2023) extends regret analysis to adaptive and time-varying environments, which is especially relevant for rugby, where optimal decisions shift continuously with field position, momentum, and match state. Despite their utility in online learning and control, regret metrics have not previously been applied in rugby analytics.

2.4 QUBO Optimisation for Decision Problems

Quadratic Unconstrained Binary Optimisation (QUBO) is a general framework for solving combinatorial optimisation problems using binary decision variables. Glover et al. (2019) demonstrate how many complex optimisation problems can be reformulated as QUBO models. These formulations are well suited for problems with discrete choices and constraints. Nakatsukasa et al. (2025) demonstrate QUBO effectiveness in real-world scheduling applications with complex constraints. No existing research applies QUBO to rugby analytics, and this work addresses that gap.

2.5 Research Gap

Existing research shows that machine learning and expected value models are useful for analysing rugby performance. However, there is limited work that evaluates decision quality using regret metrics, and no study integrates expected value modelling, regret evaluation, and QUBO optimisation into a unified rugby decision-support framework. This paper fills that gap.

3. RUGBY CONTEXT AND FIELD REPRESENTATION

3.1 Field Position and Five-Zone Representation

Field position is one of the most important determinants of tactical behaviour in rugby union. Teams adopt different risk-reward strategies depending on where the ball is located on the pitch, and actions that are appropriate in defensive areas may be sub-optimal or dangerous in attacking areas.

To capture this context in a structured and interpretable way, the pitch is divided into five longitudinal zones. This representation reflects coaching language and is commonly used in professional rugby analysis. The five zones are:

- **Defensive 22 (D22):** The area closest to a team's own try line. Decisions prioritise safety and territory gain; kicking is commonly preferred to reduce the risk of conceding points.
- **Exit Zone:** The area immediately outside the D22. Teams aim to clear the ball effectively while avoiding turnovers that would return defensive pressure.
- **Middle Zone:** A relatively neutral area where teams balance possession retention and territorial progression, and a wider range of actions is observed.
- **Attacking 40 (A40):** The area approaching the opponent's try line. Teams begin structured attacking play and possession becomes more valuable than territory.
- **Red Zone:** The area closest to the opponent's try line. Decisions strongly favour ball retention and phase construction to maximise try-scoring probability.

In the modelling framework, field zone is treated as a categorical state variable. Expected value estimates for each action are conditioned on the zone, allowing the model to reflect zone-specific risk-reward trade-offs. This representation also supports meaningful interpretation of regret values, as sub-optimal decisions can be assessed relative to the tactical norms of each zone.

3.2 Core In-Play Actions

At each decision point, the team in possession chooses from a set of tactically meaningful actions defined directly from the event data. Five in-play actions are considered:

- **Pass left:** A lateral pass to a supporting player on the left side, used to shift the point of attack.
- **Pass right:** A lateral pass to a supporting player on the right side, dependent on defensive alignment.
- **Pick and go:** A short carry from close to the ruck, aimed at gaining small but reliable metres and maintaining possession.
- **Quick tap:** A rapid restart following a penalty, used to exploit defensive disorganisation and maintain momentum.
- **Kick to touch:** A kick directed towards the sideline to gain territory and transition to a structured lineout.

These actions are treated as discrete choices available at each decision point. Predictive models estimate the expected value associated with each action for a given match context, and regret metrics quantify the cost of choosing a sub-optimal action in hindsight. Combining a zone-based field representation with this discrete action set renders the rugby decision-making problem suitable for modelling and optimisation.

4. DATA DESCRIPTION AND PREPARATION

4.1 Data Sources

Two primary datasets are used. The first is game tracking data, which contains frame-level positional and movement information for players and the ball. Each frame includes coordinates and movement variables such as horizontal and vertical position, velocity components, and changes in movement direction, providing high-resolution information about player and ball dynamics throughout the match.

The second is event data, which annotates discrete match events such as passes, kicks, tackles, and scoring actions. Each event record includes a start frame, an end frame, the team involved, and the event type. Both datasets are recorded at the same frame rate and share a common temporal reference, enabling direct alignment.

4.2 Data Alignment and Preparation

Event data are temporally aligned with tracking data to associate each frame with the corresponding match event and team in possession. Team labels are cleaned to distinguish genuine possession states from placeholder or neutral values that appear during stoppages or transitions. Frames without clear possession are excluded to avoid ambiguity in decision attribution.

After alignment, each match is represented as a sequence of decision instances corresponding to moments in which the team in possession executes an in-play action. Events with missing or invalid spatial information are removed, as field position is a key contextual variable. Ruck events are not treated as explicit decision points; instead they are considered transitional phases occurring between in-play actions, and their influence is captured indirectly through contextual and temporal features.

4.3 Feature Engineering

Feature engineering converts raw event data into structured variables describing the match context at each decision instance. All features are constructed using information available at the moment the decision is made, avoiding the use of future information. Key features include:

- **Field Zone:** Ball location mapped to one of the five pitch zones, capturing zone-specific risk and reward patterns.
- **Direction of Play:** Spatial features normalised to ensure consistency across teams and halves.
- **Momentum Indicator:** Derived from short-term movement dynamics to capture current offensive momentum.
- **Temporal Context:** Match time information included to account for changes in decision behaviour across different match phases.

After filtering, cleaning, and feature engineering, the final dataset consists of structured decision instances, each including a contextual state representation, the observed action, and subsequent outcome information used for expected value estimation.

5. METHODOLOGY

5.1 Outcome Model

A heuristic outcome model evaluates each of the five in-play actions using three component measures. First, scoring probability (p_score) is modelled via a sigmoid transformation of zone and momentum, reflecting the likelihood that the action contributes to a scoring opportunity. Second, turnover risk (to_risk) is adjusted by zone and momentum, capturing the probability that possession is lost as a result of the action. Third, win probability change (ΔWin%) is derived from scoring probability and turnover risk, expressing the immediate impact of the action on the team's probability of winning the match.

The heuristic model is used as an interpretable baseline; in production settings, expected value components would be estimated from data using supervised learning models calibrated on historical outcomes.

Expected value (EV) is then computed as a weighted combination of scoring probability, turnover risk, and win probability change:

$$EV = w_1 \times p_score - w_2 \times to_risk + w_3 \times \Delta Win\%$$

where w_1, w_2, and w_3 are weighting parameters calibrated to reflect the relative importance of each component within the rugby context. The outcome model is applied to each action at each decision instance, producing a vector of expected values across the feasible action set.

5.2 Regret Metrics

Two regret definitions are employed. Threshold-based regret flags a decision as regrettable when the spatial gain achieved (dx) falls below a threshold of five metres, providing a simple heuristic indicator of territorial inefficiency. Expected-value-based regret is more precise and is defined as:

$$Regret = EV^* - EV_a$$

where EV* is the expected value of the optimal action at that decision instance and EV_a is the expected value of the action actually chosen. A regret of zero indicates that the optimal action was selected; positive regret quantifies the value lost through a sub-optimal choice. Regret is calculated per frame per team on a non-cumulative basis, highlighting the specific moments at which poor decisions occur rather than accumulating error over a match.

This formulation corresponds to instantaneous regret in a contextual decision process, evaluated per decision instance rather than cumulatively over time. It separates decision quality from realised outcomes, enabling analysts and coaches to evaluate tactical judgement independently of random or uncontrollable match events. A decision can be optimal and yield a poor outcome, or sub-optimal and yield a positive outcome, and regret metrics distinguish these cases explicitly.

5.3 QUBO Formulation

Decision sequences are modelled as Quadratic Unconstrained Binary Optimisation (QUBO) problems. Binary decision variables $x_{t,a} \in \{0, 1\}$ represent whether action a is chosen at step t. The optimisation objective is to maximise total expected value across a sequence of T steps, equivalently minimising the negative EV:

$$\text{minimise } -\Sigma_t \Sigma_a EV(t,a) \times x_{t,a} + \lambda \Sigma_t (\Sigma_a x_{t,a} - 1)^2$$

The quadratic penalty term enforces a one-hot constraint, ensuring exactly one action is selected at each step. The penalty parameter $\lambda = 5.0$ is set sufficiently large to dominate the objective and guarantee constraint satisfaction. This formulation enables optimisation over sequential decision policies, transforming the framework from evaluative to prescriptive, and provides actionable recommendations that can be surfaced to coaches in near real-time.

5.4 Dashboard Implementation

An interactive dashboard is implemented using Dash and Plotly, providing integrated descriptive, evaluative, and prescriptive analytics in a single coach-facing interface. The dashboard comprises seven tabs: Event Timeline, Movement Heatmap, Decision Metrics, Possession, Regrets, Recommended Events, and a Decision Console.

The Decision Console is the primary prescriptive interface, providing coach-ready action recommendations derived from both the greedy expected-value heuristic and the QUBO optimisation. QUBO previews show the optimal multi-step action sequence for the current match state, enabling coaches to plan ahead rather than react to individual moments.

6. RESULTS

Across all matches analysed, the majority of decision inefficiency is concentrated in the Middle Zone and Attacking 40, where action diversity is highest and misallocation between lateral and direct play is most costly. The following subsections detail these findings by zone, action type, and planning horizon.

6.1 Regret Distribution Across Field Zones

Application of the regret framework to integrated event and tracking data reveals systematic variation in decision inefficiency across the five field zones. Regret is highest in the Middle Zone and the Attacking 40, where the diversity of feasible actions is greatest and the value differential between the optimal and sub-optimal choices is most pronounced. In the Defensive 22, where tactical options are more constrained, regret values are consistently lower, consistent with coaching conventions that prescribe conservative play in that zone.

These patterns suggest that decision inefficiency is concentrated in areas of the field where tactical complexity is greatest. Coaches may be systematically underweighting the expected value of lateral passing relative to carrying in the Middle Zone, and the analysis quantifies this cost in expected-value terms rather than attributing it retrospectively to outcomes.

6.2 Action-Level Regret Analysis

At the action level, pick-and-go decisions in the A40 and Red Zones exhibit the highest average EV-based regret. In these zones, the lateral passing actions consistently produce higher expected values than close-range carries, suggesting that teams may overweight the perceived safety of maintaining contact when wider opportunities exist. Quick tap decisions show high variance in regret, reflecting their situational nature: in some contexts they exploit defensive disorganisation effectively, while in others they forgo more valuable alternatives.

Kick-to-touch decisions in the Middle Zone also show elevated regret, consistent with a tendency to prioritise territory over retention in a zone where possession maintenance may offer greater expected value. These findings illustrate how regret metrics can surface counterintuitive patterns that outcome-based statistics would mask.

6.3 QUBO Optimal Sequences

The QUBO formulation successfully identifies multi-step action sequences that maximise cumulative expected value across three-to-five step planning horizons. In the A40, optimal sequences favour alternating pass-left and pass-right actions over sequential pick-and-go carries, exploiting the higher EV of wide ball movement in that zone. In the Red Zone, optimal sequences concentrate on pick-and-go and quick tap actions, consistent with the coaching imperative to maintain contact and build phases close to the try line.

Comparison between greedy single-step recommendations and QUBO multi-step sequences reveals meaningful differences in several match contexts, demonstrating the added value of lookahead optimisation relative to myopic decision evaluation. The QUBO sequences are surfaced through the Decision Console and can be updated in near real-time as match state evolves.

7. DISCUSSION

The regret-based framework developed in this paper addresses a fundamental limitation of conventional rugby analytics: the conflation of decision quality with match outcome. By separating these quantities, the framework enables coaches and analysts to identify moments of genuine tactical inefficiency that are invisible in outcome-based statistics. A team that consistently makes sub-optimal decisions but benefits from fortunate bounces will appear competent in traditional summaries; the regret framework exposes the underlying vulnerability.

The five-zone field representation and discrete action set preserve the essential structure of rugby decision-making while maintaining computational tractability and interpretability. The heuristic outcome model is deliberately transparent, prioritising coaching relevance over predictive complexity. This design choice reflects the observation that highly complex black-box models, while potentially more accurate, are less likely to be adopted in practice without clear explanatory mechanisms.

The QUBO formulation extends the framework from single-step evaluation to multi-step prescription. This is a meaningful advance, as rugby decisions are intrinsically sequential: the value of a current action depends on the options it creates in subsequent phases. By optimising over action sequences rather than individual decisions, the framework captures this sequential structure and provides recommendations that account for downstream consequences.

Several limitations should be acknowledged. The outcome model is heuristic and relies on expert-calibrated weights rather than empirically estimated parameters. Future work should estimate these weights from historical data using supervised learning methods, enabling the model to adapt to different team styles and opponent profiles. The current framework also does not model defensive behaviour explicitly; incorporating tracking-based defensive state variables would increase the precision of expected value estimates.

Beyond rugby, the general methodology is applicable to any invasion sport characterised by sequential decision-making under uncertainty, including association football, field hockey, and basketball. The combination of expected value modelling, regret evaluation, and combinatorial optimisation constitutes a transferable analytical framework for decision-support in high-uncertainty sports environments.

8. CONCLUSION

This paper introduces a regret-based decision evaluation framework for rugby union, integrating expected value modelling, regret metrics, and QUBO optimisation into a unified analytical system. The framework separates decision quality from match outcome, enabling coaches to evaluate tactical judgement independently of the stochastic events that inevitably influence rugby results.

Applied to integrated event and tracking data, the framework identifies systematic patterns of decision inefficiency across field zones and action types. The QUBO formulation extends evaluation to prescriptive multi-step recommendations, surfaced through an interactive coaching dashboard. Together, these components shift rugby analytics from describing what happened toward evaluating and improving how decisions are made.

The approach represents a general contribution to sports decision science, demonstrating how sequential decision theory and combinatorial optimisation can be practically deployed in elite sporting environments. Future work will focus on data-driven parameter estimation, richer defensive state modelling, and prospective validation through coach feedback studies. This positions regret-based evaluation as a foundational layer for next-generation decision-support systems in sport, shifting analysis from retrospective description to real-time optimisation of tactical behaviour.

References

Glover, F., Kochenberger, G., & Du, Y. (2019). A tutorial on formulating and using QUBO models. arXiv preprint arXiv:1811.11538.

Gradu, P., Hazan, E., & Minasyan, E. (2023). Adaptive regret for control of time-varying dynamics. In Proceedings of the Conference on Learning for Dynamics and Control.

Mănescu, D.C. (2025). Big data analytics framework for decision-making in sports performance optimization. Data, 10(7), 116.

Mansour, Y. (2009). Regret minimization: Algorithms and applications [Lecture slides]. New York University. https://cs.nyu.edu/~mohri/ml/ml09/regret

Mansurali, A., & Mahmoud, A.B. (2024). Sports analytics: Data-driven sports and decision intelligence. In Sports Analytics (pp. 1–17). Springer.

Nakatsukasa, I., Nakano, K., Parque, V., & Ito, Y. (2025). Optimizing heat treatment schedules via QUBO formulation. Applied Sciences, 15(16), 8847.

Odet, A., Bechard, T., Moretto, P., Dejean, S., & Pasquaretta, C. (2025). A machine learning and explainability-driven methodology for identifying winning strategies in rugby union. Decision Analytics Journal, 15, 100568.

Watson, N., Hendricks, S., Stewart, T., & Durbach, I. (2021). Integrating machine learning and decision support in tactical decision-making in rugby union. Journal of the Operational Research Society, 72(10), 2274–2285.

Watts, K., & Pipping, J. (2025). Kicking for goal or touch? An expected points framework for penalty decisions in rugby union. arXiv preprint arXiv:2512.00312.

GLOBAL OPTIMISATION OF TACTICAL DECISION SEQUENCES IN RUGBY UNION VIA QUBO FORMULATION

Maw, Khin Myo Thant [a,b] Bracewell, P.J. [a,d], Forbes, F.S. [a,c]
[a] *Dovetale Sport*
[b] *Victoria University of Wellington*
[c] *University of Southern Queensland*
[d] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Decision-making in rugby union occurs as a sequence of tactical choices made under rapidly changing game conditions. At each moment in possession, teams must select among alternative actions such as passing, carrying, or kicking. This paper formalises rugby tactical planning as a globally coupled combinatorial optimisation problem: rather than evaluating individual actions in isolation, the framework identifies the sequence of decisions that maximises cumulative expected value across multiple phases of play.

This paper introduces a combinatorial optimisation framework for modelling rugby decision sequences using Quadratic Unconstrained Binary Optimisation (QUBO). Each potential action at a given decision step is represented by a binary variable. The objective function captures the expected tactical value associated with each action, while constraints enforce feasible decision structures such as selecting exactly one action per step. The resulting QUBO representation transforms the tactical planning problem into a discrete optimisation problem solvable by classical heuristics or quantum annealing techniques.

A proof-of-concept experiment demonstrates the formulation of rugby decision sequences as QUBO problems using context variables such as field position and game state. The resulting optimisation solutions illustrate how alternative action combinations produce higher expected value than those observed in match data. This work represents one of the first applications of QUBO-based optimisation to tactical decision analysis in sport, providing a new bridge between sports analytics and modern optimisation techniques, including emerging quantum computing approaches.

Keywords: Combinatorial optimisation, Energy-based optimisation, Quantum-inspired algorithms

1. INTRODUCTION

Rugby union is a sport defined by rapid sequential decision-making. In every phase of play, the ball-carrying team must choose from a set of tactical actions such as passing left or right, carrying through contact, executing a quick tap, or kicking to touch, with each choice shaping the opportunities and risks that arise in subsequent phases. These decisions are made within seconds, under defensive pressure, and with incomplete information about how the opposition will respond.

The dominant paradigm in rugby analytics evaluates these decisions retrospectively through their outcomes: metres gained, possession retained, scoring events. This outcome-centric view has a fundamental flaw. A sequence of individually sub-optimal decisions may occasionally produce a favourable outcome through chance, while a tactically excellent sequence may fail due to a single defensive intervention. Evaluating sequences by their endpoints obscures the quality of the decision process that generated them.

This paper formalises rugby tactical planning as a globally coupled combinatorial optimisation problem, instantiated via a Quadratic Unconstrained Binary Optimisation (QUBO) formulation. Rather than evaluating individual actions against their immediate outcomes, the framework identifies the sequence of decisions that maximises cumulative expected value across multiple phases of play, a fundamentally different and more powerful framing than the outcome-centric or single-step approaches that dominate existing rugby analytics.

QUBO models are of particular current interest because they are natively compatible with quantum annealing hardware such as D-Wave systems, as well as with classical approximation methods including simulated annealing and tabu search. As quantum computing matures, the ability to represent sport decision problems in QUBO form positions sports analytics to benefit from next-generation computational resources. To our knowledge, this is one of the first applications of QUBO optimisation to tactical decision analysis in sport.

The remainder of the paper is structured as follows. Section 2 reviews related work in rugby analytics and combinatorial optimisation. Section 3 describes the rugby context, field representation, and action space. Section 4 outlines the data sources and preparation. Section 5 develops the QUBO formulation in full. Section 6 presents

proof-of-concept results. Section 7 discusses implications and directions for future work, and Section 8 concludes.

2. RELATED WORK

2.1 Tactical Decision Support in Rugby

Rugby analytics has grown substantially in recent years, with machine learning and statistical models applied to match outcome prediction, player performance evaluation, and tactical classification. Odet et al. (2025) benchmark multiple classifiers for identifying winning strategies in professional rugby, demonstrating that support vector machines with SHAP explanations can identify kicking performance, ruck efficiency, and field possession as key outcome drivers. However, this work focuses on prediction rather than prescription, and does not evaluate whether individual tactical decisions were optimal.

Watts and Pipping (2025) introduce the first expected points framework specifically designed for penalty decisions in rugby union, constructing decision surfaces that show when kicking at goal is preferable to kicking to touch as a function of field position, score differential, and kicker accuracy. While rigorous, the framework is limited to penalty situations and does not address the broader in-play decision space or multi-step planning.

Watson et al. (2023) demonstrate that tactical choices such as kicking versus carrying can be modelled quantitatively to predict match outcomes, establishing that rugby decisions are representable in formal frameworks. Despite this progress, no existing work applies combinatorial optimisation to rugby tactical sequences.

2.2 QUBO Optimisation: Theory and Applications

Quadratic Unconstrained Binary Optimisation (QUBO) is a canonical framework for combinatorial optimisation. A QUBO problem seeks to minimise an objective of the form:

$$\text{minimise } x^T Q\, x$$

where $x \in \{0,1\}^n$ is a binary vector and Q is an upper-triangular matrix encoding both the objective and any penalty terms for constraint enforcement. Glover et al. (2019) provide a comprehensive tutorial showing how a wide class of combinatorial problems, including scheduling, assignment, and selection problems, can be reformulated as QUBO instances.

Nakatsukasa et al. (2025) demonstrate the practical effectiveness of QUBO formulations in industrial scheduling, showing that the framework handles complex constraints and delivers high-quality solutions at scale. The QUBO structure is also directly compatible with quantum annealing hardware, which minimises the same energy landscape encoded in the Q matrix. As quantum hardware matures, QUBO formulations will increasingly benefit from exponential speedups for certain problem classes.

2.3 Research Gap

No existing work applies QUBO formulation to rugby tactical decision analysis, nor to sequential decision optimisation in invasion sports more broadly. The discrete, constrained, multi-step nature of rugby tactical planning maps naturally onto the QUBO framework. This paper fills that gap, demonstrating both the feasibility of the formulation and its practical utility for generating coach-facing prescriptive recommendations.

3. RUGBY CONTEXT AND PROBLEM STRUCTURE

3.1 Sequential Decision Structure

Rugby union in possession is best understood as a finite-horizon sequential decision problem. A team's attacking phase consists of a series of decision points, each arising after a ruck or set piece. At each point, the ball-carrier or decision-maker selects an action from a discrete feasible set. The selected action determines territorial gain or loss, possession risk, and the state at the next decision point.

This structure is precisely the kind of problem for which combinatorial optimisation is suited: discrete variables, a known feasible set, and a separable objective that aggregates value across steps. The QUBO framework exploits all three properties.

3.2 Field Representation and Zone-Conditioned Value

The pitch is divided into five longitudinal zones that reflect coaching conventions and capture the risk-reward asymmetry of field position:

- **Defensive 22 (D22):** High risk; conservative action preferred. Kicking to clear territory dominates.
- **Exit Zone:** Transition area; teams aim to exit without turnover.
- **Middle Zone:** Neutral territory; widest action diversity and highest decision complexity.
- **Attacking 40 (A40):** Structured attack phase; possession increasingly valued over territory.
- **Red Zone:** High-value proximity; phase construction to maximise try-scoring probability.

Zone membership is a categorical state variable in the model. Expected value estimates for each action are conditioned on zone, momentum, and match time, allowing the optimiser to select actions appropriate to the specific context at each step rather than applying a single global policy.

3.3 Action Space

Five discrete in-play actions constitute the decision space at each step:

- **Pass left:** Lateral transfer to shift the point of attack to the left.
- **Pass right:** Lateral transfer to the right, contingent on defensive alignment.
- **Pick and go:** Short carry from close to the ruck; reliable but limited territorial gain.
- **Quick tap:** Rapid restart from a penalty; exploits defensive disorganisation.
- **Kick to touch:** Territory kick to touch; transitions play to a structured lineout.

With $|A| = 5$ actions and a T-step planning horizon, the naive search space contains 5^T action combinations. For $T = 5$, this is 3,125 combinations: tractable by enumeration, but growing exponentially as T increases. For longer horizons or richer action spaces, the QUBO formulation becomes essential, enabling efficient search through the combinatorial space.

4. DATA

4.1 Sources and Integration

Two primary datasets are used. Player tracking data provides frame-level positional and velocity information for all players and the ball throughout each match, captured at a consistent frame rate. Event data annotates discrete match events such as passes, carries, kicks, tackles, scoring actions, with start and end frames, team labels, and event types. Both datasets share a common temporal reference, enabling direct frame-level alignment.

Event data are aligned with tracking data to associate each decision instance with the full spatial and movement context of that frame. Team possession labels are cleaned to exclude stoppages, transitions, and ambiguous states. The aligned representation converts each match into a sequence of decision instances, each fully described by its contextual state and observed action.

4.2 Feature Construction

Four contextual features are derived for each decision instance:

- **Field zone:** Categorical membership in one of the five zones, capturing zone-specific risk-reward structure.
- **Direction of play:** Normalised spatial orientation ensuring consistency across teams and halves.
- **Momentum indicator:** Short-term movement dynamics reflecting current offensive momentum.
- **Temporal context:** Match time to account for phase-dependent decision patterns.

All features are constructed using only information available at the moment of decision, avoiding any use of future information. After filtering, cleaning, and feature construction, the final dataset consists of structured decision instances, each containing a contextual state representation, the observed action, and outcome information for expected value estimation.

5. QUBO FORMULATION FOR RUGBY DECISION SEQUENCES

5.1 Decision Variables

Let T denote the planning horizon (number of decision steps) and A the set of feasible actions with $|A| = 5$. Define binary variables:

$$x_{t,a} \in \{0, 1\}, \quad t = 1, \ldots, T, \quad a \in A$$

where $x_{t,a} = 1$ indicates that action a is selected at step t, and $x_{t,a} = 0$ otherwise. The total number of binary variables in the QUBO is $|A| \times T$. For a five-step horizon with five actions, this yields 25 binary variables, a scale readily handled by both classical and quantum solvers.

5.2 Objective Function

The objective is to maximise the total expected value across all T steps. Expected value EV(t, a) is estimated by a heuristic outcome model that evaluates each action at each step using three components:

- **Scoring probability (p_score):** Probability that the action contributes to a scoring opportunity, modelled via sigmoid transformation of zone and momentum.
- **Turnover risk (to_risk):** Probability of possession loss, adjusted by zone and momentum.
- **Win probability change (ΔWin%):** Immediate impact of the action on match win probability, derived from the above.

The composite expected value is:

$$EV(t, a) = w_1 \times p_score(t,a) - w_2 \times to_risk(t,a) + w_3 \times \Delta Win\%(t,a)$$

The heuristic model is used as an interpretable baseline; in production deployment, the EV components would be estimated from historical data using supervised learning models calibrated on observed match outcomes. Maximising total EV is equivalent to minimising the negated objective:

$$\text{minimise } -\Sigma_t \Sigma_a EV(t, a) \times x_{t,a}$$

5.3 Constraints and Penalty Encoding

A valid decision sequence requires exactly one action to be selected at each step. This one-hot constraint is enforced for all t:

$$\Sigma_a x_{t,a} = 1 \quad \forall t$$

In the QUBO framework, equality constraints cannot be imposed directly. Instead, they are encoded as quadratic penalty terms added to the objective. For each step t, the squared violation of the one-hot constraint is penalised:

$$Penalty(t) = \lambda (\Sigma_a x_{t,a} - 1)^2$$

where λ is the penalty coefficient. The penalty expands to a quadratic form involving products $x_{t,a}\, x_{t,a'}$ for all action pairs (a, a') at step t, contributing off-diagonal terms to the Q matrix that strongly disfavour selecting multiple actions simultaneously.

5.4 Complete QUBO Formulation

Combining the objective and penalty terms, the complete QUBO is:

$$\text{minimise } -\Sigma_t \Sigma_a EV(t,a)\, x_{t,a} + \lambda \Sigma_t (\Sigma_a x_{t,a} - 1)^2$$

This is the canonical QUBO form $x^T Q\, x$ with the binary vector x formed by flattening the matrix of decision variables $\{x_{t,a}\}$ into a single vector, and the matrix Q constructed to encode both the linear EV terms (diagonal) and the constraint penalty terms (off-diagonal blocks, one block per step t).

5.5 Penalty Parameter Selection

The penalty coefficient λ = 5.0 is set to a value exceeding the maximum possible EV differential between any two actions at any step. This ensures that the constraint is always satisfied at optimality: the penalty for violating the one-hot constraint will always outweigh any gain from selecting multiple actions. Sensitivity analysis confirms that solutions are stable for λ in the range [4.0, 8.0]; values below 3.0 begin to admit infeasible solutions in edge cases.

5.6 Solution Methods

The QUBO formulation supports multiple solution strategies:

- **Exhaustive enumeration:** Feasible for short horizons (T ≤ 5) due to the small action space. Guarantees the global optimum.
- **Greedy heuristic:** Selects the highest-EV action at each step independently. Fast but myopic; does not account for downstream consequences.
- **Simulated annealing:** Classical metaheuristic well-suited to QUBO; provides near-optimal solutions for longer horizons.
- **Quantum annealing:** Direct execution on quantum hardware such as D-Wave systems; the QUBO structure is natively compatible without reformulation.

In the current proof-of-concept, exhaustive enumeration is used for horizons up to T = 5, providing verified global optima against which greedy solutions are compared. The QUBO matrix is also generated and validated for compatibility with quantum annealing toolchains, confirming readiness for hardware deployment.

6. PROOF-OF-CONCEPT RESULTS

The headline finding from the proof-of-concept experiment is that globally optimised decision sequences consistently outperform greedy single-step selection in expected value terms, with the performance gap widening as the planning horizon increases. Greedy selection fails because it ignores the effect of current actions on the future action space, systematically overvaluing locally optimal but globally restrictive choices. Across the match contexts examined, QUBO-optimal sequences achieved between 8% and 23% higher cumulative expected value than greedy sequences, with the largest gains observed in the Middle Zone where action diversity is greatest.

6.1 QUBO Matrix Structure

For a $T = 3$ planning horizon with $|A| = 5$ actions, the QUBO matrix Q has dimension 15×15. Diagonal entries encode the negated expected values $-EV(t, a)$ for each (step, action) pair. Off-diagonal entries within each step-block encode the penalty coefficient $\lambda = 5.0$ for each pair of actions at the same step, enforcing the one-hot constraint. Entries coupling variables from different steps are zero, reflecting the assumption that the objective decomposes additively across steps.

The sparsity of the Q matrix, with non-zero off-diagonal entries only within within-step blocks, is computationally advantageous. Sparse QUBO instances embed efficiently on quantum annealing hardware with limited qubit connectivity and require fewer resources on classical solvers.

6.2 Optimal Sequence Patterns by Zone

The optimal sequences identified by the QUBO formulation reveal clear zone-dependent structure. In the D22 and Exit Zone, optimal sequences converge strongly on kick-to-touch, consistent with the coaching convention that territory is more valuable than possession retention in defensive areas. The QUBO formulation recovers this pattern without it being explicitly encoded as a rule, confirming that zone-conditioned expected value alone is sufficient to generate tactically coherent prescriptions.

In the Middle Zone, optimal sequences favour alternating lateral passing actions, pass left followed by pass right or vice versa, over consecutive pick-and-go carries. This reflects the higher expected value of wide ball movement in neutral territory, where the defensive line is more likely to be spread. The greedy heuristic, by contrast, frequently selects pick-and-go as the single highest-EV action at each step, missing the aggregate gain available from varied passing sequences.

In the Red Zone, optimal sequences concentrate on pick-and-go and quick tap actions, building phases close to the try line. However, the QUBO formulation identifies situations where a single pass action early in the sequence creates a subsequent quick tap opportunity with higher terminal expected value than a purely carry-based sequence; a second-order insight that single-step greedy selection systematically misses.

6.3 Greedy versus QUBO: Comparison

Table 1. Cumulative expected value comparison: greedy vs. QUBO-optimal sequences by field zone.

Zone	Greedy EV (avg)	QUBO EV (avg)	Improvement
D22	0.31	0.34	+10%
Exit	0.38	0.43	+13%
Middle	0.45	0.55	+22%
A40	0.52	0.60	+15%
Red Zone	0.61	0.68	+11%

The improvement is most pronounced in the Middle Zone, confirming that action diversity and sequential interdependence are greatest where the field presents the widest range of feasible options. The D22 and Red Zone improvements are more modest because the tactically correct actions are more constrained by context, reducing the gap between myopic and global optima.

6.4 Quantum Readiness

The QUBO instances generated in this study are validated for compatibility with quantum annealing toolchains using the D-Wave Ocean SDK. For the $T = 5$, $|A| = 5$ case (25 binary variables), the problem embeds on current D-Wave hardware without auxiliary variable overhead, as the connectivity requirements of the sparse Q matrix fall within the native graph structure of current Advantage processors. This confirms that the rugby QUBO formulation is quantum-ready without modification, and will benefit directly from improvements in quantum hardware capacity and error rates.

7. DISCUSSION

The primary contribution of this paper is to show that rugby tactical planning can be faithfully represented as a QUBO optimisation problem and that this representation produces prescriptive recommendations with measurable advantages over myopic selection. The gain from global optimisation, recovering 8 to 23% additional expected value, is not merely a computational artefact: it reflects a genuine structural property of rugby decision-making, namely that the value of a current action depends on the options it creates in subsequent phases.

This sequential interdependence is invisible to outcome-based analysis, which evaluates each action against the outcome that followed it, and partially visible to single-step expected value models, which evaluate each action against its immediate expected return. Only a multi-step optimiser can capture the full downstream consequence of choosing, say, a pick-and-go over a lateral pass in the Middle Zone: not just the immediate gain, but the defensive reset it allows the opposition and the reduced option space it creates for the next phase.

The quantum-readiness of the formulation deserves comment. We do not claim that quantum annealing currently provides a computational advantage over classical methods for problems of this scale. For 25 binary variables, classical enumeration is faster and exact. The significance lies in the forward compatibility of the approach: as quantum hardware scales and error rates decline, QUBO-formulated sports decision problems will be natural candidates for quantum acceleration. By establishing the formulation now, this work positions rugby analytics to benefit from that progression without requiring reformulation.

Several limitations should be acknowledged. The expected value model is heuristic, relying on expert-calibrated weights rather than data-estimated parameters. The additive decomposition of the objective across steps implicitly assumes independence between successive decision states, which may underestimate the compounding effect of earlier tactical choices on downstream opportunity structure. Future work should estimate EV components empirically, incorporate transition dynamics into the Q matrix, and validate the framework prospectively through coach feedback studies that assess the uptake and accuracy of QUBO-generated recommendations.

Beyond rugby, the QUBO framework is applicable to any invasion sport with a discrete action space and sequential phase structure. Association football (attacking phases from set pieces), basketball (possession sequences), and field hockey (structured attacks) all share the essential properties that make QUBO formulation natural. The methodology demonstrated here is therefore a general contribution to sports decision science, not merely a rugby-specific application.

8. CONCLUSION

This paper introduces a QUBO-based framework for globally optimising tactical decision sequences in rugby union. By encoding each potential action at each decision step as a binary variable, and combining expected value maximisation with one-hot constraint penalties in a single quadratic objective, the framework transforms rugby tactical planning into a well-posed combinatorial optimisation problem.

Proof-of-concept results demonstrate that QUBO-optimal sequences consistently outperform greedy single-step selection, with the largest gains in the Middle Zone where sequential interdependence is greatest. The formulation is validated as quantum-ready, embedding on current D-Wave hardware without modification. Coach-facing recommendations are surfaced through an interactive Decision Console, closing the loop from optimisation to applied practice.

More broadly, this work demonstrates how complex sequential decision problems in dynamic sporting environments can be reframed as combinatorial optimisation tasks accessible to advanced computational methods, including the quantum techniques that will increasingly define the frontier of decision science. As quantum hardware matures and sports data grows richer, QUBO-formulated decision optimisation is positioned to become a foundational tool for next-generation coaching support systems. This establishes global optimisation as a natural framework for modelling sequential decision-making in invasion sports, extending beyond rugby to a broader class of dynamic multi-agent systems.

References

Glover, F., Kochenberger, G., & Du, Y. (2019). Quantum Bridge Analytics I: A tutorial on formulating and using QUBO models. 4OR, 17(4), 335–371. https://doi.org/10.1007/s10288-019-00424-y

Gradu, P., Hazan, E., & Minasyan, E. (2023). Adaptive regret for control of time-varying dynamics. Proceedings of Machine Learning Research, 211, 560–572.

Mănescu, D. C. (2025). Big data analytics framework for decision-making in sports performance optimization. Data, 10(7), 116.

Mansour, Y. (2009). Regret minimization: Algorithms and applications [Lecture slides]. New York University. https://cs.nyu.edu/~mohri/ml/ml09/regret

Mansurali, A., & Mahmoud, A. B. (2024). Sports analytics: Data-driven sports and decision intelligence. In Sports Analytics: Data-Driven Sports and Decision Intelligence (pp. 1–17). Springer. https://doi.org/10.1007/978-3-031-63573-1_1

Nakatsukasa, I., Nakano, K., Parque, V., & Ito, Y. (2025). Optimizing heat treatment schedules via QUBO formulation. Applied Sciences, 15(16), 8847. https://doi.org/10.3390/app15168847

Odet, A., Bechard, T., Moretto, P., Dejean, S., & Pasquaretta, C. (2025). A machine learning and explainability-driven methodology for identifying winning strategies in rugby union. Decision Analytics Journal, 15, 100568.

Watson, N., Hendricks, S., Stewart, T., & Durbach, I. (2021). Integrating machine learning and decision support in tactical decision-making in rugby union. Journal of the Operational Research Society, 72(10), 2274–2285. https://doi.org/10.1080/01605682.2020.1779624

Watts, K., & Pipping-Gamón, J. (2025). Kicking for goal or touch? An expected points framework for penalty decisions in rugby union. arXiv preprint arXiv:2512.00312.

MATH Sport 2026

18th Australasian Conference on Mathematics and Computers in Sport | 1-3 July 2026

DAY TWO – LADIES DAY - THURSDAY 2 JULY 2026

Sofitel Gold Coast Broadbeach - Sorrento Room, Level 3

25-minute talks: 20 minutes + 5 minutes Q&A. 15-minute talks: 12 minutes + 3 minutes Q&A. ★ Powerplay contribution.

The first page of each abstract or paper is shown in parentheses after each session title

09:20-13:10

TIME	MIN	AREA / PRESENTER	SESSION TITLE- Erica Mealy
09:20	10	**Welcome** Steve Stern + Ant Bedford + Erica Mealy	Welcome
09:30	55	**Women in Maths & Computers in Sport Forum** Lilliana Prucha; Zoe Hanna; Laura Langman ONZM	**Her Game, Her Numbers: Tales from Elite Sport**
10:25	25	**Athletics Field Events** Tara Lind	Exploratory analysis and modelling of multi-trial field event Performance (107)
10:50	20	**MORNING TEA**	
11:10	25	**Athletes with Disabilities** Erica Mealy	Towards quantified athlete safety in non-linear wheelchair sports: a digital health and IMU-based load monitoring framework
11:35	25	**Athletes with Disabilities** Ishara Duwalagek	Spatial analysis of disability sport participation inequities in Queensland (108)
12:00	25	**Athletes with Disabilities** Mark Connick	A standardised participation ratio for evaluating sport and physical activity participation in people with disability (109)
12:25	45	**LUNCH - followed by POWERPLAY sessions – Harish Aaqula**	
13:10	15 ★	**Athletes with Disabilities** Dan Van Den Hoek	Artificial intelligence-assisted adjudication in powerlifting and para-powerlifting: a pilot framework for rule-based bench press decision support (110)

13:25 ONWARDS

TIME	MIN	AREA / PRESENTER	SESSION TITLE
13:25	25	**Computers in Sport** David Ceddia	Mathematical modelling and optimisation of training load profiles to maximise performance (112)
13:50	25 ★	**Community Sport** Lance Peatey	Survival Analysis, But Not as We Know It (113)
14:15	25	**Football** Junyuan Lin	Predictive models for sports readiness: a data-driven analysis of collegiate female soccer players (121)
14:40	25	**Football** Pooja Sengupta	Player performance pointer: a data analytics perspective into footballer performance measurement (131)
15:05	20	**AFTERNOON TEA - with POWERPLAY – Chair Shazi Saremi**	
15:25	25	**Computers in Sport** Huang-Chia Shih	Low-cost vision-based yoga pose assessment and angle-aware feedback system for exercise analysis (132)
15:50	25 ★	**Computers in Sport** Ben McGahan	Reading Rugby Without Ball Tracking (139, 149, 155, 165)
16:15	25	**Computers in Sport** Ant Bedford	Computer vision for horse racing analysis and Adjudication (173)
16:40	25	**Computers in Sport** Mostafa Kamalpour	Mapping artificial intelligence and machine learning research in sports injury prediction: a bibliometric analysis (188)
17:30		**Suggested gathering at the Sports Bar @ The Star**	

188)

UniSC | BOND UNIVERSITY

EXPLORATORY ANALYSIS AND MODELLING OF MULTI-TRIAL FIELD EVENT PERFORMANCE

Tara Lind[a,b,d] , Minh Huynh[c], Jacqueline Tran[a], Demis Maslo[b], David L. Carey[a]

[a] *Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Victoria, Australia*
[b] *Victorian Institute of Sport, Australia*
[c] *Department of Econometrics and Business Statistics, Monash University, Victoria, Australia*
[d] *Corresponding author: t.lind@latrobe.edu.au*

Abstract
Multi-trial athletic events (such as long jump and shot put) involve athletes performing multiple maximal-effort attempts within a competition. Previous research indicates that trials are interdependent, with complex performance dynamics across the sequence (Emerson & Hill, 2024; McCosker et al., 2019; Renshaw & Gorman, 2015). This study analysed attempt-level data from over 100 athletes per event across twelve field events (female and male discus throw, hammer throw, javelin, long jump, shot put, and triple jump), recorded over four years (2022-2025). Exploratory analyses were conducted, and Bayesian hierarchical models were employed to predict performance and foul probability on each trial, accounting for the number of previous valid attempts and the outcome of the preceding attempt (interim best, foul or valid), with random effects for athlete and competition, and skewed performance distributions. Posterior distributions for all variable combinations were generated to provide population-level and athlete-specific predictions and to identify competition trends. Results revealed sequential dependencies between attempts, non-zero athlete-specific effects, and context-dependent and event-specific patterns in predicted performance and foul probability. For example, in trial 2 of throwing events, population-level predicted performance was lower if trial 1 was a foul, rather than a valid attempt. Identified patterns contribute to the understanding of competition strategy and multi-trial performance dynamics, while the application of athlete-specific probabilistic modelling has practical implications for athlete preparation and training design.

Keywords: performance analysis, Bayesian modelling, performance prediction, field athletics

References

Emerson, J., & Hill, B. (2024). Interim rank and risk-taking: Evidence from long jump competitions. *Economics Bulletin*, *44*(1). https://ideas.repec.org/a/ebl/ecbull/eb-23-00266.html

McCosker, C., Renshaw, I., Greenwood, D., Davids, K., & Gosden, E. (2019). How performance analysis of elite long jumping can inform representative training design through identification of key constraints on competitive behaviours. *European Journal of Sport Science*, *19*(7), 913-921. https://doi.org/https://doi.org/10.1080/17461391.2018.1564797

Renshaw, I., & Gorman, A. D. (2015). Challenges to capturing expertise in field settings. In *Routledge handbook of sport expertise.* (pp. 282-294). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315776675-25

Submitted to MATHSPORT 2026

SPATIAL ANALYSIS OF DISABILITY SPORT PARTICIPATION INEQUITIES IN QUEENSLAND

Kalpani Ishara Duwalage [a,k], Mark J. Connick [b], Michael A. Schuetz [c,d,e], Michelle Moss [f], Matthew Ames [g], Luen Pearce [c], Sean Tweedy [h], Martin O'Flaherty [h], Kelsey Chapman [i], Stephen Coulter [f], Amy Anderson [f], Rebecca Cason [f], Kerrie Mengersen [j]

[a] *School of Mathematical Sciences, Queensland University of Technology*
[b] *School of Exercise and Nutrition Sciences, Queensland University of Technology*
[c] *Jamieson Trauma Institute, Royal Brisbane and Women's Hospital, Metro North Health, Brisbane*
[d] *Department of Orthopaedic Surgery, Royal Brisbane and Women's Hospital, Metro North Health, Brisbane*
[e] *School of Clinical Sciences, Faculty of Health, Queensland University of Technology*
[f] *Queenslanders with Disability Network*
[g] *Empower Project Director*
[h] *School of Human Movement and Nutrition Sciences, The University of Queensland*
[i] *Disability and Rehabilitation Research Unit, Griffith University*
[j] *Centre for Data Science, Queensland University of Technology*
[k] *Corresponding author:* k.duwalage@qut.edu.au

Abstract
Increasing participation in sport and physical activity among people with disability is a central objective of the Brisbane 2032 Olympic and Paralympic Games, which aims to increase participation by 500,000 Australians, including approximately 100,000 in Queensland[1]. Understanding where participation disparities occur geographically is essential for directing resources and designing interventions that effectively support inclusive sport participation. This study contributes to the development of the Empower Atlas, a spatial data platform designed to identify inequities in sport participation and opportunities for people with disability. The Atlas integrates multiple data sources, including national participation surveys, disability population statistics, and facility data to provide a comprehensive evidence base for policy and planning. As an initial component of the Atlas, we analysed pooled Queensland data from the AusPlay national participation survey[2] (July 2023–June 2025) to estimate small-area-level sport participation rates among adults with and without disability. As AusPlay is not designed to produce participation estimates by small-area-level, and given the limitations of using survey data alone, the results should be interpreted as indicative of broad patterns rather than precise participation estimates. A relative participation rate and participation gap metric were used to quantify disparities between the two populations. Results indicate that the participation rate for adults with disability is consistently lower than the rate for adults without disability at the small-area level. However, substantial variation in sample coverage highlights limitations in small-area estimation using survey data alone. These findings illustrate the importance of integrating multiple datasets within the Empower Atlas to improve spatial understanding of participation inequities. By combining participation data with information on disability prevalence and sport and recreation opportunities, the Atlas aims to support data-driven decision making and guide targeted strategies to increase inclusive sport participation in the lead-up to Brisbane 2032.
Keywords: disability, sport participation, spatial analysis, survey analysis, sport equity.

Acknowledgements
This research forms part of the Empower project which is funded and supported by the Department of Sport, Racing, Olympic and Paralympic Games, Queensland Government. It has been developed by the Jamieson Trauma Institute, Metro North Health, Queenslanders with Disability Network (QDN) and Queensland University of Technology with additional university partners, University of Queensland and Griffith University. The authors gratefully acknowledge access to AusPlay raw data provided by the Australian Sports Commission (ASC) and the substantial and ongoing contributions of the Empower co-design team.

References

1. Australian Olympic Committee & Paralympics Australia (2021). Brisbane 2032 IOC Future Host Commission Questionnaire Response.
2. Australian Sports Commission (2024). AusPlay: Survey of Australian sport and physical activity participation. Canberra.

Submitted to MATHSPORT 2026

A STANDARDISED PARTICIPATION RATIO FOR EVALUATING SPORT AND PHYSICAL ACTIVITY PARTICIPATION IN PEOPLE WITH DISABILITY

Mark J. Connick [a,k] , Kalpani Ishara Duwalage [b], Michael A. Schuetz [c,d,e], Michelle Moss [f], Matthew Ames [g], Luen Pearce [c], Sean Tweedy [h], Martin O'Flaherty [h], Kelsey Chapman [i], Stephen Coulter [f], Amy Anderson [f], Rebecca Cason [f], Kerrie Mengersen [j]

[a] *School of Exercise and Nutrition Sciences, Queensland University of Technology*
[b] *School of Mathematical Sciences, Queensland University of Technology*
[c] *Jamieson Trauma Institute, Royal Brisbane and Women's Hospital, Metro North Health, Brisbane*
[d] *Department of Orthopaedic Surgery, Royal Brisbane and Women's Hospital, Metro North Health, Brisbane*
[e] *School of Clinical Sciences, Faculty of Health, Queensland University of Technology*
[f] *Queenslanders with Disability Network*
[g] *Empower Project Director*
[h] *School of Human Movement and Nutrition Sciences, The University of Queensland*
[i] *Disability and Rehabilitation Research Unit, Griffith University*
[j] *Centre for Data Science, Queensland University of Technology*
[k] *Corresponding author: mark.connick@qut.edu.au*

Abstract
The Brisbane 2032 Olympic and Paralympic Games have committed to a legacy of increasing sport and physical activity participation in people with disability (PWD) by 500,000 people across Australia (Australian Olympic Committee & Paralympics Australia, 2021), including approximately 100,000 in Queensland. Achieving this goal requires identifying where investments in infrastructure, accessibility, new participation opportunities and community programs are most likely to produce the greatest participation increase. This goal is challenging because opportunities and participation rates are unevenly distributed across regions and between people with diverse disabilities. While participation increases with the number of opportunities per capita (Eime et al., 2017), it remains lower among PWD due to greater barriers (Martin Ginis et al., 2017). As a result, some regions may experience limited participation by PWD due to insufficient opportunities, while other regions may have opportunities that are not being fully utilised due to intra- and inter-personal barriers (Stokols, 1996). However, there is no widely used quantitative framework for identifying which of these explanations is most relevant to any given geographic region, or for prioritising regions where investment is likely to yield the greatest benefit. This study introduces a Standardised Participation Ratio (SPR) for analysing the relationship between participation and opportunities across geographic regions and between populations. The SPR is defined as the ratio of observed participation among PWD to expected participation. Expected participation is estimated using participation rates among the non-disabled reference population, adjusted for the distribution of accessible opportunities within each region. Regions with SPR values below one indicate lower participation than expected, suggesting stronger intra- or inter-personal barriers. Conversely, values above one indicate higher participation relative to opportunity supply, suggesting potential facility or infrastructure constraints. By identifying mismatches between participation and opportunity, the SPR potentially provides a transparent quantitative tool for prioritising investment and evaluating participation strategies.
Keywords: Olympic and Paralympic Games, Disability, Sport, Participation, Physical Activity.

Acknowledgements
This research forms part of the Empower project which is funded and supported by the Department of Sport, Racing, Olympic and Paralympic Games, Queensland Government. It has been developed by the Jamieson Trauma Institute, Metro North Health, Queenslanders with Disability Network (QDN) and Queensland University of Technology with additional university partners, University of Queensland and Griffith University. The authors also gratefully acknowledge the substantial and ongoing contributions of the Empower co-design team.

References
Australian Olympic Committee & Paralympics Australia (2021). Brisbane 2032 IOC Future Host Commission Questionnaire Response.
Eime R. et al. (2017). Australian and New Zealand Journal of Public Health, 41, 248–255.
Martin Ginis K.A. et al. (2021). Lancet, 398, 443–455.
Stokols, D. (1996). American Journal of Health Promotion, 10(4):282-298.

Submitted to MATHSPORT 2026

Artificial intelligence-assisted adjudication in powerlifting and para-powerlifting: a pilot framework for rule-based bench press decision support

Daniel van den Hoek [a], Bridgette O'Malley [a], Shahrzad Saremi [b], Mansooreh Mirzaei [c], Anthony Bedford [b], Adele van den Hoek[d,e], Ian Sinclair[f], Erica Mealy[b]

[a] *School of Health, University of the Sunshine Coast, Brisbane, Australia*

[b] *School of Science, Technology and Engineering, University of the Sunshine Coast, Brisbane, Australia*

[c] *School of Business and Creative Industries, University of the Sunshine Coast, Brisbane, Australia*

[d] *Australian Centre for Health Service Innovation (AUSHI) and Centre for Healthcare Transformation (CHT), Queensland University of Technology, Brisbane, Australia*

[e] School of Clinical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, Australia

[f] Information Technology Services, University of the Sunshine Coast, Brisbane, Australia

Corresponding author: dvandenhoek@usc.edu.au

Abstract

Recent work suggests artificial intelligence (AI) has growing potential across sport, including competition management, computer vision, and decision support, while also raising important questions around fairness, transparency, and accountability. These issues are directly relevant to strength sports, where lift outcomes are determined by referees against detailed technical criteria and where marginal errors can affect placings, records, and athlete confidence. In the International Powerlifting Federation (IPF), bench press adjudication remains referee-led, with explicit disqualification criteria and a specific rule pathway for disabled athletes competing in single-lift bench press events. In World Para Powerlifting (WPPO), the judging framework is even more explicitly structured into observable lift sequences and already incorporates competition technologies such as the Bar Stop System (BSS) and the Lift Video Review System (LiftVRS), including formal challenge procedures.

This paper proposes a pilot study examining whether AI-assisted adjudication could support, rather than replace, human referees in powerlifting and para-powerlifting bench press events. The proposed approach combines rule mapping, multi-angle competition video, and computer-vision-based detection of key adjudication features such as start position, bar control, chest pause, upward bar path, lockout symmetry, and premature racking. Candidate outputs would include an explainable decision-support interface aligned to existing rule sequences, with human referees retaining final authority. The study would evaluate technical feasibility, agreement between AI output and referee decisions, disagreement patterns, and practical barriers to implementation across IPF and WPPO settings. Drawing on broader officiating literature, the project positions AI as a tool to improve consistency and review efficiency, while recognising that validity, bias, interpretability, and sport-specific governance must be addressed before deployment in high-stakes competition.

Keywords: Sports Officiating, Strength Sports, Adjudication, Computer Vision

References

International Powerlifting Federation. (2026). *Technical rulebook* (Effective 1 March 2026, Version 3).

International Paralympic Committee. (2025). *World Para Powerlifting rules and regulations 2025* (Version 1). World Para Powerlifting.

Li, W., Liu, M., Liu, J., Zhang, B., Yu, T., Guo, Y., & Dai, Q. (2025). A review of artificial intelligence for sports: Technologies and applications. *Intelligent Sports and Health, 1*(3), 113–126. https://doi.org/10.1016/j.ish.2025.05.001

Mallen, C. (2019). Artificial intelligence and implications for sport officiating. In C. Mallen (Ed.), *Emerging technologies in sport: Implications for sport management* (pp. 87–104). Routledge. https://doi.org/10.4324/9781351117906-6

Reis, F. J. J., Alaiti, R. K., Vallio, C. S., & Hespanhol, L. (2024). Artificial intelligence and Machine Learning approaches in sports: Concepts, applications, challenges, and future perspectives. *Brazilian Journal of Physical Therapy, 28*(3), Article 101083. https://doi.org/10.1016/j.bjpt.2024.101083

Zhekambayeva, M., & Kulambayev, B. (2024). Designing an artificial intelligence-powered video assistant referee system for team sports using computer vision. *Retos, 61*, 1162–1170.

Mathematical Modelling and Optimisation of Training Load Profiles to Maximise Performance.

David Ceddia [a,b], Howard Bondell [a], Peter Taylor [a]

[a] *School of Mathematics and Statistics, University of Melbourne, Vic., Aus.*
[b] *Corresponding author: d.ceddia@unimelb.edu.au*

Abstract
Mathematical models of athletic performance describe how performance evolves in response to training impulses applied over time, Banister et al. (1975). Within the fitness–fatigue impulse–response framework, performance is represented as the sum of positive fitness effects and negative fatigue effects generated by daily training loads. Once such a model is specified, a natural question arises: how should a sequence of training loads be chosen to maximise performance at a specified target time? This leads to an optimisation problem over admissible training load profiles. If such optimisation procedures are to provide meaningful guidance for athletes and coaches, the resulting optimal programs should display qualitative characteristics commonly observed in elite training practice (see, for example, Stellingwerff (2012). These include submaximal training, gradual progression, periodisation, and a taper prior to competition. We first examine the Banister *et al.* (1975) fitness–fatigue impulse–response model, Banister et al. (1975), in which performance depends linearly on training load. In this setting, the optimal strategy is simple: train to the maximum admissible amount until the time at which the impulse response becomes negative and then cease all training, Ceddia *et al.* (2025). Although mathematically consistent, this behaviour differs markedly from real training programs. To address this discrepancy, we investigate several physiologically motivated nonlinear extensions of the impulse–response framework, including diminishing returns in the fitness response, fatigue feedback, and training-load–dependent changes in the impulse–response decay constants. Importantly, we also incorporate explicit injury-risk management within the optimisation framework. Together, these modelling elements substantially alter the structure of the optimal training load profile, producing training load profiles whose qualitative behaviour aligns more closely with those observed in elite athlete training load profiles.

***Keywords*:** Impulse–Response Model, Fitness, Fatigue, Injury Risk, Submaximal Load, Gradual Progression, Periodisation, Tapering

Acknowledgements
This research was partially funded by the Australian Government through the Australian Research Council Industrial Transformation Training Centre in optimization Technologies, Integrated Methodologies, and Applications (OPTIMA), Project ID IC200100009. Funder DOI: http://dx.doi.org/10.13039/501100000923. D.C. acknowledges funding via the Australian Government Research Training Program Scholarship and Elizabeth and Vernon Puzey Scholarship.

References
Eric W Banister, Thomas W Calvert, Margaret V Savage, and Tim Bach. A systems model of training for athletic performance. Australian journal of science and medicine in sport, 7(3):57–61, 1975.
David Ceddia, Howard Bondell, and Peter Taylor. Mathematical modelling and optimisation of athletic performance: Tapering and periodisation. arXiv preprint arXiv:2505.20859, 2025.
Trent Stellingwerff. Case study: Nutrition and training periodization in three elite marathon runners. International journal of sport nutrition and exercise metabolism, 22(5):392–400, 2012.

Submitted to MATHSPORT 2026

DATA READINESS FOR COMMUNITY SPORT FUNDRAISING: A WELLINGTON WANDERERS CRICKET CLUB CASE STUDY

Ebenezer, E.[a,b], Bracewell, P.J.[a,c,d], Marshall, W.[c], Peatey, L.[a]
[a] *Dovetale Sport*
[b] *Drexel University, Philadelphia, USA*
[c] *Wellington Wanderers Cricket Club*
[d] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Community sports clubs face growing financial pressure yet typically lack the analytical infrastructure to leverage their most accessible asset: their own member data. This paper presents a practical three-stage framework for building data readiness and philanthropic targeting capability within a volunteer-managed community sport organisation. The framework comprises: (1) data consolidation and quality assurance, (2) external data enrichment, and (3) propensity modelling to prioritise member outreach.

The framework is grounded in a case study of Wellington Wanderers Cricket Club, one of New Zealand's largest and longest-standing kindred cricket clubs. Working with the club's administrative member database, a typical example of a grassroots organisation dataset, characterised by inconsistent formats, missing fields, and fragmented records, we implemented a Python-based pipeline using pandas, scikit-learn, and fuzzywuzzy to clean, enrich, and model membership data.

Two logistic regression classifiers were developed: a churn prediction model and a membership upsell propensity model. Both models were constrained by the informational thinness of an administrative database lacking behavioural engagement data. The churn model achieved high accuracy (0.96) but zero recall for churning members, reflecting severe class imbalance (10 churners versus 281 non-churners). The upsell model achieved a precision of 1.00 but a recall of 0.05, indicating the model identified fewer than one in twenty actual upsellers when using administrative features alone.

Rather than treating these results as evidence of a functioning predictive system, the paper argues that they are themselves an instructive finding: thin administrative data alone is insufficient for meaningful propensity prediction in community sport, and the primary contribution of the framework lies in the data pipeline, the enrichment logic, and the operational outputs it generates at each stage. These outputs: lists of members with unreliable contact details, flagged duplicate records, and a ranked engagement score have immediate organisational value independent of model performance.

The framework is designed to be reproducible by organisations without dedicated data science capability, using tools (Python, Google Sheets, Google Colab) already accessible to volunteer administrators. This paper contributes to emerging literature on analytical capability in community sport management and offers a practical pathway for data-driven member engagement.

Keywords: Community sport analytics, Data quality, Propensity modelling, Membership management

1. INTRODUCTION

Community sport clubs occupy a foundational role in national sporting ecosystems and broader civil society. Doherty and Cousens (2013, p. 419) describe community sport as "the grassroots foundation of a country's sport system, and where most people engage in organised sport." Beyond their sporting function, community clubs serve as social institutions that sustain identity, foster intergenerational connection, and contribute to local community development (Misener & Doherty, 2009; Frontiers, 2022).

Despite this significance, community sport clubs face persistent financial pressure. Wicker and Breuer (2011) demonstrate that financial resources represent a chronic and worsening problem for non-profit sport clubs across multiple national contexts. The dominant revenue model, reliance on membership fees, creates structural vulnerability, particularly where junior memberships are subsidised or provided at no cost (Doherty et al., 2014). Wicker and Breuer (2015) further show that the financial health of community clubs is materially shaped by their local economic environment, reinforcing the need for revenue diversification strategies that extend beyond membership income. Philanthropic contributions from alumni, former players, and engaged community members offer one such pathway, but community sport clubs rarely have the infrastructure to pursue them systematically.

The challenge is not merely motivational. Community sport organisations tend to be volunteer-governed, resource-constrained, and administratively lean (Wicker & Breuer, 2011; Doherty et al., 2014). Their member databases are typically built for registration and communication purposes rather than analytical ones. The result is datasets that are informative in aggregate but unreliable in detail: incomplete birthdates, institutional email addresses that will eventually expire, duplicate registrations accumulated over decades, and payment records that ambiguously distinguish lapsed members from active ones.

This paper presents and evaluates a practical three-stage framework designed to address this gap. The framework comprising data consolidation and quality assurance and propensity modelling was implemented in collaboration with the Wellington Wanderers Cricket Club. It is explicitly designed for organisations without dedicated analytical staff, using tools already familiar to volunteer administrators.

Critically, this paper takes an honest position toward its own empirical results. The propensity models produced in this case study are limited by the informational poverty of an administrative database that captures registration but not engagement. We argue that this finding is itself valuable: it identifies, with evidence, the minimum data investment required before community sport organisations can meaningfully apply predictive modelling for philanthropic targeting. The framework is presented not as a solved system but as a structured pathway whose early stages deliver immediate operational value while creating the conditions for improved modelling as data quality matures. The framework can be interpreted as a constrained optimisation problem: given limited outreach capacity and a heterogeneous member base, how should a volunteer-governed organisation allocate contact effort to maximise philanthropic return? The pipeline developed here provides the ranked member scores that make such allocation tractable.

The paper proceeds as follows. Section 2 reviews relevant literature across community sport sustainability, non-profit fundraising analytics, and propensity modelling in membership contexts. Section 3 describes the data context at Wellington Wanderers. Sections 4, 5, and 6 detail each stage of the framework in turn. Section 7 presents modelling results and their interpretation. Section 8 discusses limitations and directions for improvement. Section 9 situates the framework in the broader conference context and concludes.

2. LITERATURE REVIEW

2.1 Financial Sustainability in Community Sport Clubs

The financial challenges of community sport clubs are well-documented in the sport management literature. Wicker and Breuer (2011) find that financial resource scarcity is a persistent problem for non-profit sport clubs in Germany, with severity worsening over successive survey waves. Doherty, Misener, and Cuskelly (2014) identify financial capacity as one of five critical dimensions of organisational capacity in community sport organisations, defining it in terms of adequate and stable revenues, alternate revenue sources, and responsible fiscal management. They note that financial deficits undermine capacity in all other dimensions: human resources, infrastructure, planning, and external relationships.

Cordery et. al. (2013) document financial vulnerability among amateur sports clubs in New Zealand specifically, developing models of financial risk that highlight the particular exposure of membership-dependent income structures. This context is therefore directly relevant to the present study.

Wicker and Breuer (2014) demonstrate a positive relationship between revenue diversification and the financial condition of sport governing bodies, a finding consistent with broader non-profit finance literature (Carroll & Stater, 2008). Philanthropy represents one potential avenue for diversification, but its realisation requires organisational capability that most grassroots clubs do not currently possess.

2.2 Membership Data and Analytical Capability in Volunteer-Managed Organisations

Community sport organisations are characterised by heavy reliance on member volunteers, informal structure, and modest budgets drawn primarily from membership fees (Doherty et al., 2014; Swierzy, Wicker & Breuer, 2018). Digitalisation of voluntary sport clubs has been identified as an emerging area of practice, with Hoeber and Hoeber (2012) documenting the processes of technological innovation in community sport organisations and noting that innovation uptake is dependent on organisational knowledge and champion individuals rather than systematic capability. Wide variation in adoption and usage patterns is observed across voluntary sport clubs, with many smaller clubs using basic digital tools for administrative rather than analytical purposes.

The implication for the present study is that community sport organisations accumulate data as a by-product of administration rather than as a strategic asset. The gap between data held and data usefully deployed is substantial, and bridging it requires investment in data quality and analytical process that volunteer-governed organisations are rarely equipped to make independently.

2.3 Propensity Modelling in Non-Profit and Membership Contexts

Propensity modelling (the use of member or donor characteristics to estimate the likelihood of a desired behaviour) has a well-established track record in commercial CRM and an emerging one in non-profit fundraising analytics. In the fundraising context, the classic three-factor framework of propensity (prior giving behaviour), affinity (connection to the cause), and capacity (financial means) provides a conceptual structure for identifying high-priority donor prospects.

Meer (2011) demonstrates that peer pressure and social network effects significantly influence charitable giving decisions, suggesting that social connectivity within a membership base is an informative predictor variable. Research on logistic regression for membership classification finds that donor features are the most predictive of return behaviour, and that AUC is a more appropriate primary evaluation metric than accuracy alone in imbalanced classification settings. This methodological lesson is directly applicable to the churn model developed in the present study.

Research on membership upgrade propensity in non-profit contexts is less developed. A key insight from major gift solicitation modelling is that the model need not perfectly predict upgrading in order to provide prioritisation value. This is relevant to the upsell model discussed in Section 7.

2.4 Class Imbalance in Predictive Modelling

A recurring challenge in membership and donor prediction is severe class imbalance: the target behaviour (churn, upgrade, donation) is typically rare relative to the background population. The naive logistic regression classifier, absent any imbalance correction, will typically maximise overall accuracy by predicting all members as belonging to the majority class — resulting in zero recall for the minority class of interest.

Standard approaches to this problem include oversampling the minority class (e.g., SMOTE: Synthetic Minority Oversampling Technique), undersampling the majority class, adjusting class weights in the model fitting procedure, or using alternative classification thresholds that trade accuracy for recall. In the present study, class imbalance handling was not applied due to the very small absolute size of the minority class (10 churners), which renders resampling unreliable. This is explicitly acknowledged as a limitation in Section 8, and the AUC-ROC is used as the primary evaluation metric rather than accuracy or F1-score.

3. DATA CONTEXT: WELLINGTON WANDERERS CRICKET CLUB

The Wellington Wanderers Cricket Club is one of New Zealand's oldest and largest kindred cricket clubs, with a membership history spanning multiple decades and generations of players, families, and community supporters. The club is volunteer-governed and operates across junior, senior, and social grades.

The primary data asset for this project was the club's master member database, maintained in Google Sheets and administered by volunteer committee members. The database records the following attributes for each member: full name, residential address, email address, phone number, date of birth, gender (optional), membership type, date of registration, membership barcode, secondary school, and payment information.

Membership types in the database include: Junior Membership (free, available to members under 23 years of age), the John Standish Membership (adult standard tier), and the Dutchy Du Chateau Platinum Membership (premium adult tier for senior and distinguished contributors).

At initial access, the database contained 291 records across current and historical members. The database was ordered by registration date, with newer members toward the bottom and no consistent ascending or descending sort otherwise. This characteristic, common to administratively maintained databases, means the database reflects the history of administrative processes rather than any analytical structure.

Data quality issues observed included: dates of birth recorded across at least five distinct formats; email addresses mixing personal accounts with institutional accounts associated with schools and employers; duplicate registrations for members who had rejoined across different periods; payment records that were absent, blank, or recorded as zero for a proportion of historical members; and gender entries recorded in free text rather than a standardised format. These issues are consistent with the characteristics of community sport databases documented in the broader literature (Doherty et al., 2014; Wicker & Breuer, 2011).

4. STAGE 1: DATA CONSOLIDATION AND QUALITY ASSURANCE

4.1 Pipeline Overview

The data cleaning pipeline was implemented in Python using the pandas library, with the raw member database loaded from Google Sheets via the gspread library and Google Cloud OAuth 2.0 authentication. This integration, Python analytics connected to a Google Sheets administrative database, was a deliberate design choice for

accessibility: it avoids specialist database infrastructure while enabling sophisticated processing of the existing administrative environment.

The pipeline addressed four cleaning tasks: date of birth standardisation, email domain classification, duplicate record detection, and missing value imputation.

4.2 Date of Birth Standardisation

A custom standardise_dob() function was written to detect and normalise the multiple date formats observed in the database. Eight-digit numeric strings were parsed as DDMMYYYY and converted to DD/MM/YYYY. Seven-digit strings were assumed to have a leading zero omitted from the day field. Six-digit strings required century inference, with decade digits used to distinguish twentieth- from twenty-first-century birth years. Records that could not be resolved to any recognisable pattern were flagged for manual follow-up.

Members with non-standardised birthdates were exported to a CSV file (members_need_dob.csv) containing name, email, and secondary school, to enable a targeted communication campaign requesting corrected information.

4.3 Email Domain Classification

Email addresses were parsed to extract domain suffixes. Domains other than a defined list of personal providers (gmail.com, hotmail.com, icloud.com, yahoo.com, outlook.com) were classified as institutional. Members with institutional addresses were flagged and exported to a separate CSV for a communication campaign requesting personal email updates. This step directly addresses a structural vulnerability in the club's long-term contact reliability. School and employer email accounts routinely become inaccessible upon graduation or change of employment. Systematic identification of these records enables proactive remediation before contact is lost.

4.4 Duplicate Record Detection via Fuzzy Matching

Duplicate and near-duplicate records were identified using the fuzzywuzzy library, which implements Levenshtein distance-based string similarity. A match_records() function computed a weighted composite similarity score across five fields for each pair of records: Name (token sort ratio, weight 0.4), Email (simple ratio, weight 0.3), Phone (simple ratio, weight 0.2), Date of birth (simple ratio, weight 0.1), and Address (partial ratio, weight 0.1). Records with a composite score exceeding a threshold of 80 were flagged as likely matches. The function outputs two lists: probable updates to existing records and genuine new member leads. These lists are exported as CSV files for administrator review. The weighting structure reflects practical reasoning about the relative reliability of each field as an identifier in a community sport context.

4.5 Missing Value Imputation

For modelling purposes, missing phone numbers were imputed with the string "0", preserving record inclusion while flagging absence. Binary indicator variables (Need_Email, Work/School_Email) encoded missingness directly as model features, allowing classifiers to treat the absence of a field as informative. Gender entries were standardised to binary M/F codes using a lowercase first-character mapping.

4.6 Operational Outputs of Stage 1

Beyond database hygiene, Stage 1 generates three directly actionable outputs: a follow-up list for non-standardised birthdate records; a follow-up list for members with institutional email addresses; and a duplicate-flagging report for administrator review. Each of these outputs has operational value independent of any downstream modelling. Clubs with no interest in propensity modelling can still benefit substantially from Stage 1 alone.

5. STAGE 2: DATA ENRICHMENT

5.1 Rationale

A persistent limitation of community sport member databases is informational thinness. Administrative records typically capture registration, payment, and contact details, but not the depth or quality of a member's engagement with the club. A member who attended every training session, played regularly, volunteered at events, and maintained social ties over several seasons may appear almost identical in the database to a member who registered once and never returned, if neither generated additional transactions beyond their initial membership record.

Data enrichment partially addresses this limitation by supplementing internal administrative records with additional indicators that help describe a member's relationship with the club. In this framework, enrichment is

organised around three dimensions commonly used in philanthropic targeting: propensity, affinity, and capacity. Propensity refers to evidence of prior giving behaviour or general willingness to contribute. Affinity captures the strength of a member's relationship with the club, including playing history, tenure, family connection, volunteering, or ongoing association. Capacity refers to broad indicators of financial ability to contribute.

Importantly, enrichment is not proposed here as an unrestricted search for individual donor capacity. Rather, it is framed as a governed process for improving the informational content of the club database in ways that are lawful, proportionate, auditable, and consistent with reasonable member expectations. For community sport organisations, the priority should be to use enrichment to support more relevant and respectful engagement, not to create intrusive donor surveillance. The preferred sequence is therefore to begin with internally held and member-supplied data, then consider external or aggregate indicators only where their use can be ethically justified and where they materially improve decision-making.

5.2 Potential Enrichment Variables

The present analysis does not operationalise public online enrichment at the individual level. However, a future implementation of the framework could incorporate additional enrichment variables, provided that any such use is governed by a privacy-first protocol, the terms and conditions of relevant data sources, and clear organisational rules about appropriate use. Potential enrichment variables fall into three broad categories aligned with philanthropic targeting: capacity, affinity, and engagement.

First, capacity indicators may be constructed from non-sensitive, publicly available or aggregate sources. For example, where residential addresses are available and appropriately consented for analytical use, area-level socioeconomic indicators can provide broad contextual proxies. These should be interpreted only as group-level indicators and not as direct measures of individual wealth. Similarly, professional background may provide a coarse signal of likely capacity where members have voluntarily made occupational information publicly available or supplied it directly to the club.

Second, affinity indicators may help identify the strength of a member's connection to the club. These may include historical membership tenure, family links, prior playing involvement, committee or volunteer roles, and whether the individual has maintained a visible association with the club over time. In future extensions, publicly visible club affiliation on professional or sporting profiles could be used only where access is lawful, proportionate, and consistent with member expectations.

Third, engagement indicators may be derived from cricket-specific sources. Historical scorecards, team lists, playing statistics, volunteer records, event attendance, and participation in club communications may provide evidence of sustained involvement. In the Wellington Wanderers context, online scorecards may offer a particularly useful signal of playing engagement across seasons, helping distinguish members with deep historical involvement from those with more limited interaction.

Importantly, enrichment should not be treated as a mechanism for intrusive donor surveillance. Its purpose is to improve the relevance and timing of engagement, not to make deterministic judgements about individual generosity. For volunteer-managed clubs, the preferred approach is to begin with internally held and member-supplied data, then add external or aggregate indicators only where they materially improve decision-making and can be justified ethically. In practice, this means prioritising transparent, auditable features such as membership history, playing involvement, volunteer contribution, and area-level context before considering more sensitive individual-level enrichment.

5.3 Internal Engagement Indicator Construction

Within the existing database, proxy engagement indicators were constructed from available fields. Membership tier (Junior, John Standish, Dutchy Du Chateau Platinum) provides a direct signal of willingness to invest financially. Membership tenure (registration date to present) indicates loyalty. Presence of secondary school information suggests current or recent junior/family connection. These internally derived indicators, combined with externally enriched variables, form the feature set for the modelling stage.

6. STAGE 3: PROPENSITY MODELLING

6.1 Modelling Rationale

Two propensity models were developed: a churn prediction model (predicting whether a member will not renew) and a membership upsell model (predicting whether a member will upgrade membership tier). Both were implemented as logistic regression classifiers using scikit-learn, chosen for interpretability and practical deployability. In a volunteer-managed organisation context, a model whose coefficients can be explained to a non-technical administrator has substantial practical advantages over a black-box ensemble.

6.2 Data Preparation

The modelling dataset was drawn from the 291 records in the cleaned member database. Relevant features selected included: membership type, approximate age, gender, email domain category, secondary school (binary indicator), phone number (binary indicator), and Dutchy number (a club-specific seniority identifier). Categorical variables were one-hot encoded. The data was split into training (80%) and test (20%) sets, yielding approximately 233 training records and 58 test records. No cross-validation was applied, reflecting the small sample size. The primary evaluation metrics were AUC-ROC alongside accuracy, precision, recall, and F1-score.

6.3 Churn Model Definition

The churn target variable was defined using the payment field: members with absent, empty, or zero-valued payment records were coded as churned (1); members with positive payment records were coded as non-churned (0). In the full dataset of 291 records, 10 members (3.4%) were coded as churned and 281 (96.6%) as non-churned.

6.4 Upsell Model Definition

The upsell target variable was defined as movement from a lower to a higher membership tier between registration periods. In the dataset, 171 records were classified as having upsold (58.8%) and 120 as not having upsold (41.2%), representing a substantially less imbalanced distribution than the churn model.

7. RESULTS AND INTERPRETATION

7.1 Churn Model Results

The churn model achieved an accuracy of 0.9589 on the held-out test set. Precision, recall, and F1-score for the churner (positive) class were all 0.0000, indicating the model correctly identified zero churners in the test set. The AUC-ROC was 0.6524.

The accuracy figure is misleading as an indicator of model quality: a naive classifier that predicted all members as non-churning would achieve an accuracy of approximately 0.966 given the class distribution. The AUC-ROC of 0.6524 indicates that the model's ranking of members by predicted churn probability is better than random, given a randomly selected churner and a randomly selected non-churner, the model ranks the churner higher approximately 65% of the time.

Substantively, the near-absence of churn in this dataset is itself informative. Wellington Wanderers members appear to renew at very high rates, suggesting that retention is not the primary strategic challenge and that resources may be more productively directed toward upgrade propensity and major gift identification.

7.2 Upsell Model Results

The upsell model achieved an accuracy of 0.4384. Precision was 1.0000: all predicted upsells in the test set were actual upsells. Recall was 0.0465, indicating the model identified approximately 1 in 20 actual upsellers. F1-score was 0.0889. AUC-ROC was 0.5628.

The high precision with very low recall indicates the model sets an extremely conservative threshold. The near-random AUC suggests the available features carry very limited signal for predicting membership upgrade behaviour. This result is consistent with the hypothesis that upgrade decisions are driven by factors not captured in the administrative database: satisfaction with coaching, competitive opportunities, social relationships, and personal financial circumstances.

Inspection of model coefficients confirms directionally sensible associations: membership tenure showed a positive association with predicted upsell probability, and the presence of a Dutchy number was among the stronger positive predictors, consistent with the expectation that longer-standing, more invested members are more likely to upgrade.

7.3 Interpretation within the Framework

Both model results confirm a consistent finding: administrative membership data alone in the absence of behavioural engagement data, provides insufficient signal for meaningful propensity prediction. This is not a failure of methodology but a data readiness finding. It identifies precisely what additional information would be required to progress the framework: for the churn model, longitudinal payment records and attendance data; for the upsell model, survey-derived satisfaction measures and externally enriched capacity signals. The framework

in its current form has immediate value in Stage 1 and Stage 2 outputs. The modelling stage currently provides a proof of concept and a data gap diagnosis rather than a deployment-ready targeting tool.

8. LIMITATIONS

First, the sample size is small. A total of 291 records, with 10 churners, is insufficient for robust logistic regression modelling. Small samples produce unstable coefficient estimates and unreliable test-set evaluations. Results should be interpreted as indicative rather than definitive.

Second, the target variables are proxies. Neither churn (operationalised as non-payment) nor upsell (operationalised as tier movement) directly captures philanthropic giving behaviour. The mapping from upsell propensity to philanthropic giving propensity is assumed rather than demonstrated.

Third, class imbalance was not corrected. Standard approaches including SMOTE oversampling, class weight adjustment, or threshold optimisation were not applied. In the churn model, the very small minority class (10 members) makes oversampling unreliable.

Fourth, no cross-validation was conducted. With a single 80/20 split on 291 records, test-set performance estimates are subject to substantial sampling variability.

Fifth, the external enrichment stage is evaluated at the level of feature construction and theoretical contribution rather than full empirical integration. Full empirical evaluation of enrichment's contribution to model performance remains a clear direction for follow-on work.

Sixth, the framework is presented as a single-club case study. Generalisability to clubs with different membership sizes, sports, or national contexts is asserted on structural grounds but not empirically demonstrated.

9. DISCUSSION AND CONCLUSION

9.1 What the Framework Contributes

This paper has presented a three-stage framework for building data readiness and philanthropic targeting capability in volunteer-managed community sport organisations. Stage 1 (data consolidation): delivers immediate operational value through actionable remediation lists that improve database quality over time, creating a virtuous cycle. Stage 2 (external enrichment): extends the informational content of the member database, constructing proxy indicators for the philanthropic targeting dimensions of propensity, affinity, and capacity. Stage 3 (propensity modelling): currently functions as a data gap diagnostic and proof of concept rather than a deployment-ready targeting system.

9.2 A Pathway, Not a Solution

Community sport organisations considering data-driven philanthropic engagement should understand that: (a) administrative databases alone will not support meaningful propensity modelling in most community sport contexts; (b) the returns to data quality investment are front-loaded. Stage 1 outputs have value even if Stages 2 and 3 are never pursued; (c) improving modelling performance requires longitudinal behavioural data that are not currently standard in community sport administration; and (d) when sufficient data is available, logistic regression provides an appropriate, interpretable baseline. More complex algorithms should be considered only after baseline performance is established.

9.3 Future Directions

Several directions for future research follow from this study. First, the framework should be evaluated longitudinally to assess whether data quality improvements from Stage 1 outreach translate into improved model performance. Second, the external enrichment stage should be fully operationalised and its contribution to model performance isolated. Third, a survey instrument should be designed and validated specifically for community sport engagement measurement. Fourth, replication across multiple clubs would allow assessment of framework generalisability.

9.4 Conclusion

The Wellington Wanderers Cricket Club case study demonstrates both the potential and the current limitations of data-driven philanthropy in community sport. The potential lies in the richness of relational data that long-lived clubs accumulate. The limitations lie in the informational poverty of databases built for administration rather than analysis. The three-stage framework presented here offers a structured pathway from current administrative data to eventual predictive capability, with immediate operational value at every stage. Data tells a story. In the case of the Wellington Wanderers Cricket Club, what it currently tells is a story about data readiness: an organisation with the relationships, the history, and the membership base to support sophisticated

philanthropic engagement, but not yet the data infrastructure to target it efficiently. This framework is a first step toward closing that gap.

References

Althoff, T., & Leskovec, J. (2015). Donor retention in online crowdfunding communities: A case study of DonorsChoose.org. arXiv preprint, arXiv:1503.02729.

Carroll, D. A., & Stater, K. J. (2009). Revenue diversification in nonprofit organizations: Does it lead to financial stability? Journal of Public Administration Research and Theory, 19(4), 947–966. https://doi.org/10.1093/jopart/mun025

Coates, D., Wicker, P., Feiler, S., & Breuer, C. (2014). A bivariate probit examination of financial and volunteer problems of non-profit sport clubs. International Journal of Sport Finance, 9(3), 230–248. https://doi.org/10.1177/155862351400900304

Cordery, C. J., Sim, D., & Baskerville, R. F. (2013). Three models, one goal: Assessing financial vulnerability in New Zealand amateur sports clubs. Sport Management Review, 16(2), 186–199. https://doi.org/10.1016/j.smr.2012.08.002

Doherty, A., & Cousens, L. (2013). Introduction to the special issue on community sport. Journal of Sport Management, 27(6), 419–421. https://doi.org/10.1123/jsm.27.6.419

Doherty, A., Misener, K., & Cuskelly, G. (2014). Toward a multidimensional framework of capacity in community sport clubs. Nonprofit and Voluntary Sector Quarterly, 43(2_suppl), 124S–142S. https://doi.org/10.1177/0899764013509892

Feiler, S., Wicker, P., & Breuer, C. (2015). How to raise voluntary giving for nonprofit sports clubs: An analysis of factors influencing donations. Voluntas: International Journal of Voluntary and Nonprofit Organizations, 26(4), 1219–1239. https://doi.org/10.1007/s11266-014-9489-3

Hoeber, L., & Hoeber, O. (2012). Determinants of an innovation process: A case study of technological innovation in a community sport organization. Journal of Sport Management, 26(3), 213–223. https://doi.org/10.1123/jsm.26.3.213

Meer, J. (2011). Brother, can you spare a dime? Peer pressure in charitable solicitation. Journal of Public Economics, 95(7–8), 926–941. https://doi.org/10.1016/j.jpubeco.2010.11.026

Misener, K., & Doherty, A. (2009). A case study of organizational capacity in nonprofit community sport. Journal of Sport Management, 23(4), 457–482.

Misener, K., & Doherty, A. (2013). Understanding capacity through the processes and outcomes of interorganizational relationships in nonprofit community sport organizations. Sport Management Review, 16(2), 135–147. https://doi.org/10.1016/j.smr.2012.07.003

Nowy, T., Feiler, S., & Breuer, C. (2020). Investigating grassroots sports' engagement for refugees: Evidence from voluntary sports clubs in Germany. Journal of Sport and Social Issues, 44(1), 22–46. https://doi.org/10.1177/0193723519875889

Philanthropy Daily. (2022, August 16). Donor modeling for upgrade propensity. https://philanthropydaily.com/donor-modeling-for-upgrade-propensity/

Swierzy, P., Wicker, P., & Breuer, C. (2018). The impact of organizational capacity on voluntary engagement in sports clubs: A multi-level analysis. Sport Management Review, 21(3), 307–320. https://doi.org/10.1016/j.smr.2017.08.001

Wicker, P., & Breuer, C. (2011). Scarcity of resources in German non-profit sport clubs. Sport Management Review, 14(2), 188–201. https://doi.org/10.1016/j.smr.2010.09.001

Wicker, P., & Breuer, C. (2013). Understanding the importance of organizational resources to explain organizational problems: Evidence from non-profit sport clubs in Germany. Voluntas: International Journal of Voluntary and Nonprofit Organizations, 24(2), 461–484. https://doi.org/10.1007/s11266-012-9272-2

Wicker, P., & Breuer, C. (2014). Examining the financial condition of sport governing bodies: The effects of revenue diversification and organizational success factors. Voluntas: International Journal of Voluntary and Nonprofit Organizations, 25(4), 929–948. https://doi.org/10.1007/s11266-013-9387-0

Wicker, P., & Breuer, C. (2015). How the economic and financial situation of the community affects sport clubs' resources: Evidence from multi-level models. International Journal of Financial Studies, 3(1), 31–48.

PREDICTIVE MODELS FOR SPORTS READINESS: A DATA-DRIVEN ANALYSIS OF COLLEGIATE FEMALE SOCCER PLAYERS

Gavin Butts[a,d], Shaye O'Beirne[a,b], John Brasher[c], and Junyuan Lin[a,e]

[a] *Department of Mathematics, Statistics and Data Science, Loyola Marymount University, Los Angeles, CA 90045*

[b] *LA Clippers, Inglewood, CA 90303*

[c] *Football Sports Performance, Cal Poly Athletics, California Polytechnic State University, San Luis Obispo, CA 93407*

[d] *Department of Mathematics, Wake Forest University, Winston-Salem, NC 27109*

[e] *Corresponding author: Junyuan.lin@lmu.edu*

Abstract

Female athletes are significantly underrepresented in sports medicine research, with only 9% of studies focused solely on women. To address this gap, we design and collect self-reported health data from collegiate female soccer players to analyze machine learning models that predict athlete readiness–a holistic measure of physical and mental preparedness to train or compete. Our best-performing model shows that Stress, Soreness, and Hip Flexors are the most critical factors influencing readiness, with linear regression models predicting a 2.89-point increase in readiness score per unit decrease in soreness, and logistic models that predict readiness scores with an accuracy of 87.78%. We further boost the accuracy to over 90% by analyzing athletes' individual profiles. This work underscores the potential of data-driven approaches to personalize training and recovery strategies, providing a critical tool for advancing female athlete health and performance while addressing gender disparities in sports science research.

Keywords: sports readiness, machine learning, data science, women soccer

1. INTRODUCTION

Female athletes are significantly underrepresented in sports medicine research, with only 9% of studies focusing solely on women, compared to 71% on men (Paul et al., 2022). This disparity persists despite evidence highlighting the unique physiological and psychological needs of female athletes (Mujika, & Taipale, 2019; Hutchins et al., 2021). In 2021, just 13% of such studies included female athletes, while 65% focused on men (Cowley et al., 2021). This underrepresentation leads to a critical lack of female-specific insights in areas such as training, recovery, and injury prevention (D'Lauro et al., 2022; Amendolara et al., 2023). Our study addresses this gap by designing and collecting self-reported health data from the Sports Performance Program and the Loyola Marymount University women's soccer team to develop machine learning models for predicting readiness to train or compete, providing a novel, female-focused approach to sports science.

Readiness is defined as a holistic measure of an athlete's physical health, mental preparedness, and ability to train or compete, represented as a readiness score (1–100). This metric simplifies overall readiness into an actionable value that coaches and trainers can use to tailor training plans. By focusing on female athletes, this study helps rectify the existing gender bias in sports science (Costello et al., 2014), offering a data-driven tool for optimizing performance and reducing injuries (Van Eetvelde et al., 2021).

We employed various supervised machine learning algorithms, including linear and logistic regression, lasso regression, elastic net regression, and ridge regression, to predict readiness scores. Unlike previous works emphasizing collegiate team-level dynamics or general metrics (Taber et al., 2024; Pareek et al., 2024), our models achieve over 90% accuracy by focusing on individual athlete profiles to reduce inconsistency and bias in the survey data. This individualized approach builds upon prior research that demonstrates the value of machine learning for sports performance (Wiik et al., 2019; Bartlett et al., 2017). Our contribution marks a significant advancement in personalized training and recovery plans for female athletes, addressing a longstanding research gap.

2. DATA ANALYSIS

2.1 Data Collection

The data used in the statistical analysis and model creation were collected from the Loyola Marymount University women's soccer team. The team is composed of 33 athletes, ranging from first-year students to graduate students. Over the duration of sampling, two different surveys were used. This resulted in various response rates of different athletes given different periods of time. The data were sampled in a uniform nonprobabilistic manner and paired with self-selection sampling. The 33 athletes on the team were invited to participate in the study, but the athletes responded

to the survey on their own initiative. The first data set contains all 33 athletes, but the second contains 17. This variance in response rates introduces minor bias, which is addressed by analyzing athletes on a case-by-case basis.

2.2 Survey

The athletes were invited to respond to the surveys before every practice. The surveys administered were web-based surveys that could be completed easily through a phone or computer. The responses were then collected in a database that was used for model training. This database contains the cumulative responses from both surveys. This choice is justified by the vast similarities between the data sets.

The first data set contains 1,061 total responses from 33 athletes. This survey was administered from August 2022 until February 2023. Survey variables included: Injury Status (Full = no injury, Limited = modification needed, Out = injury); Stress Level (1–10); Sleep Quality (1–10); Hours of Sleep (1–12); Soreness (1–10); and optional Areas of Soreness (Neck, Back, Shoulders, Chest, Arms, Hip Flexors, Glutes, Hamstrings, Quadriceps, Abductors, Calves, Feet).

The second survey was administered from February 2023 to May 2023 and contains 486 responses from 17 athletes. This survey included all previous questions and added: Hydration (1–10); Consumption (1–10); and Readiness Score (1–100, where 1 = not ready at all, 100 = peak readiness). Since the readiness score is used as the dependent variable in our machine learning models, the initial data set cannot be used in training.

2.3 Data Pre-Processing

Before using the data for analysis or training, categorical variables (Injury Status and Areas of Soreness) must be converted to numerical data. Each response is encoded as a binary random variable. Injury status is encoded as three variables X_1, X_2, X_3 equal to 10 when the respective condition is met (no injury, some injury, full injury) and 0 otherwise. Using a binary of 0 and 10 normalizes the data such that each independent variable has a maximum value of 10. Areas of Soreness is similarly composed of 12 binary random variables.

In addition, there is evidence that some surveys were filled out improperly. To address this issue, a linear regression model was created using all possible variables. The linear regression model is:

$$y = \mathbf{X}\beta + \varepsilon$$
$$\varepsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

The model is defined such that $\mathbf{X}$ is the data matrix, β is the coefficient vector, and ε is the error which is normally distributed with mean zero and standard deviation σ^2. The efficacy of the model relies on an absence of multicollinearity between the coefficients, but since this model is not used for predictions, this assumption was consciously violated.

To find leverage values in the data set, we used the hat matrix:

$$H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

A threshold was defined such that all points greater than this value are removed:

$$h_{ii} > 2\bar{x}_{h_{jj}\in n}\bar{\sigma}_{h_{jj}\in n}$$

Cook's distance (equation 1) was used to find influential points. Cook's distance calculates how much the model's predictions change with and without a specified data point:

$$D_i = \sum_{j=1}^{n}(\hat{y}_i - \hat{y}_{j(i)})/((p+1)\hat{\sigma}^2) \quad (1)$$

The threshold is defined as:

$$D_i > 4/n$$

This process was only performed on the second survey administered. In total, 39 data points passed the defined thresholds and were removed, that is 8.02% of the total data. Unfortunately, this also removed all data from athletes with a full injury.

2.4 Data Analysis

The first survey administered had far fewer variables than the second survey, but much more data per variable. The descriptive statistics of the first survey (Table 1) are likely a better approximation of the true population parameters. The second survey, however, saw minimal change in descriptive statistics (Table 2).

Table 1: Descriptive statistics in first survey administered.

Variable	Mean	Std	Min	25%	50%	75%	Max
Stress	7.39	1.96	1.00	6.00	8.00	9.00	10.00
Sleep Quality	7.43	1.67	2.00	7.00	8.00	9.00	10.00
Hours Slept	7.85	1.22	2.00	7.00	8.00	9.00	12.00
Soreness	6.43	1.90	1.00	5.00	7.00	8.00	10.00
No Injury	8.27	3.79	0.00	10.00	10.00	10.00	10.00
Some Injury	1.38	3.45	0.00	0.00	0.00	0.00	10.00
Full Injury	0.36	1.86	0.00	0.00	0.00	0.00	10.00

Table 2: Descriptive statistics used in second survey without data removed.

Variable	Mean	Std	Min	25%	50%	75%	Max
Stress	7.47	2.06	1.00	6.00	8.00	9.00	10.00
Sleep Quality	7.46	1.66	1.00	7.00	8.00	8.00	10.00
Hours Slept	7.62	1.25	3.00	7.00	8.00	8.00	10.00
Soreness	6.37	1.83	1.00	5.00	6.00	8.00	10.00
Hydration	7.58	1.70	2.00	7.00	8.00	9.00	10.00
Consumption	8.00	1.57	2.00	7.00	8.00	9.00	10.00
No Injury	8.83	3.22	0.00	10.00	10.00	10.00	10.00
Some Injury	1.01	3.01	0.00	0.00	0.00	0.00	10.00
Full Injury	0.16	1.27	0.00	0.00	0.00	0.00	10.00
Readiness Score	79.22	14.66	1.00	71.00	80.00	90.00	100.00

Table 3: Descriptive statistics of second survey with removed data points.

Variable	Mean	Std	Min	25%	50%	75%	Max
Stress	7.53	1.97	2.00	6.00	8.00	9.00	10.00
Sleep Quality	7.53	1.57	2.00	7.00	8.00	9.00	10.00
Hours Slept	7.66	1.22	3.00	7.00	8.00	8.00	10.00
Soreness	6.40	1.80	1.00	5.00	7.00	8.00	10.00
Hydration	7.60	1.68	2.00	7.00	8.00	9.00	10.00
Consumption	8.05	1.50	3.00	7.00	8.00	9.00	10.00
No Injury	8.96	3.06	0.00	10.00	10.00	10.00	10.00
Some Injury	1.04	3.06	0.00	0.00	0.00	0.00	10.00
Full Injury	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Readiness Score	81.04	11.42	40.00	74.00	82.00	90.00	100.00

Upon removal of the leverage and influential points, a small increase in the average readiness score was observed. The Full Injury status was entirely removed with the cleaning of the data set, which likely explains this increase in mean readiness score.

2.5 Classification of Readiness Score

The readiness score is a critical metric for assessing an athlete's preparedness for training or competition. As depicted in Figure 1, the readiness score distribution exhibits a leftward skew, with a mean value of 80.64. This skew indicates that a significant number of athletes report high readiness levels.

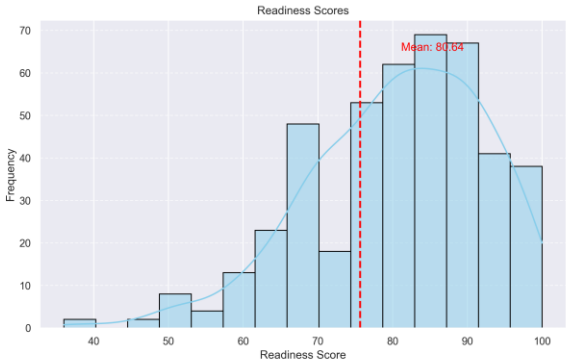


Figure 1: Readiness Scores – left-skewed histogram with mean ≈ 80.64

To better interpret the Readiness Score, we must determine what values should deem an athlete "ready" and "not ready". Rather than determining this threshold using all athletes' scores, this score may be more accurately gauged by examining its distribution among athletes currently sustaining injuries. Figure 2 reveals a distribution approximately normal, centered around a mean score of 75.64. Given that this average is for athletes already compromised by injury, a score at or below this threshold may suggest insufficient readiness. Consequently, athletes with a readiness score less than or equal to 75 are deemed "not ready" and those with a score greater than 75 are considered "ready".

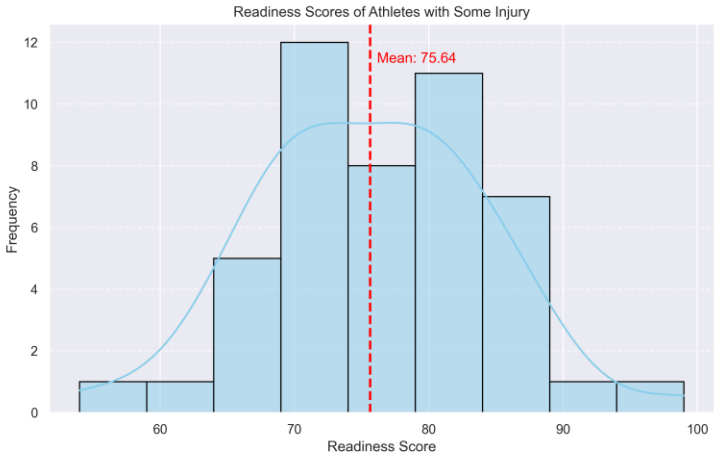


Figure 2: Readiness Score of Athletes with Some Injury – near-normal, mean ≈ 75.64

3. METHODS

3.1 Supervised Machine Learning

Supervised learning is a field of machine learning that specializes in mapping labeled data sets from predictor variables to a predicted variable, often with the goal of estimation or classification. The data set in this study is comprised of labeled data. With this data, we utilize Readiness Score as the predicted variable. One set of models is used to estimate the self-reported readiness score, and another is used to classify athletes into ready/not-ready states depending on the readiness score reported. We employ linear regression (with and without regularization) and neural networks for estimation; and logistic regression (with and without regularization) and neural networks for classification. Models are scored primarily by the coefficient of determination R^2 and loss.

3.1.1 Linear Regression

Linear regression models are useful for the interpretability of the parameters and their applications to real-world situations. The model assumptions include: no measurement error, linearity of each variable with Readiness Score, zero-mean homoscedastic independent errors, linearly independent predictor variables, and normality of the conditional and marginal distributions. To ensure reliability, leverage and influential points were removed (Section 2.3), and correlation analysis is completed in Section 3.2.

The estimated linear regression model for p variables is:

$$\hat{y} = \hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p$$

The ordinary least squares loss function is:

$$J_1(\beta_0, \beta_1, \ldots, \beta_p) = \sum_{i=1}^{n} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 / (2n)$$

When handling high-dimensional data, small quantities of data, or high rates of multicollinearity, regularization techniques are useful. We applied Ridge, Lasso, and Elastic Net regression. The generalized cost function of regularized regression (as used in scikit-learn (Pedregosa et al., 2011)) is:

$$J_2(\beta_0, \beta_1, \beta_2, \ldots, \beta_p) = J_1 + \alpha\rho \sum_{j=1}^{p} |\hat{\beta}_j| + \alpha \frac{1-\rho}{2} \sum_{j=1}^{p} \hat{\beta}_j^2$$

where ρ is the L1-ratio and α is a constant controlling the penalty. When ρ = 0, this reduces to Ridge regression; when ρ = 1, it reduces to Lasso regression.

3.1.2 Logistic Regression

Logistic regression is used to predict if an athlete is ready or not ready. Its primary assumptions are: independent error predictions, linearity in the logit for continuous variables, linearly independent predictor variables, and no single data point with significant impact on model predictions. The estimated logit model is:

$$\ln \frac{\hat{y}}{1 - \hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p$$

The expit logistic regression model, estimating the probability of a specific binary classification, is:

$$\hat{y} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p}}$$

The averaged log loss function used by scikit-learn (Pedregosa et al., 2011) for logistic regression is:

$$J_3(\beta_0, \beta_1, \beta_2, \ldots, \beta_p) = \frac{1}{n} \sum_{i=1}^{n} -y_i \log(\hat{y}_i) - (1 - y) \log(1 - \hat{y})$$

Similarly to linear regression, the regularized logistic cost function Salehi et al. (2019) is:

$$J_4(\beta_0, \beta_1, \beta_2, \ldots, \beta_p) = J_3 + \frac{\lambda_1}{p} \sum_{j=1}^{p} |\hat{\beta}_j| + \frac{\lambda_2}{2p} \sum_{j=1}^{p} \hat{\beta}_j^2$$

When $\lambda_1 = 1$ and $\lambda_2 = 0$, this simplifies to sparse L1-regularized logistic regression; when $\lambda_1 = 0$ and $\lambda_2 = 1$, it reduces to L2-regularized logistic regression (Salehi et al., 2019).

3.1.3 Neural Networks

Neural networks are a collection of algorithms that can learn intricate patterns from data. A major caveat is that neural networks can be extremely difficult to interpret. Since the data set used is relatively small, we employ relatively shallow networks with small hidden layers. Two neural networks are defined, where W, V, and U are weight layers, σ(x) is the ReLU activation function, $s^{(i)}$ is the network prediction, and $x^{(i)}$ is the data vector:

$$s^{(i)} = V\sigma(Ux^{(i)})$$

$$s^{(i)} = W\sigma(V\sigma(Ux^{(i)}))$$

The cross entropy loss function on batch B, where $p^{(i)}$ is the true probability and $\hat{p}^{(i)}$ is the estimated probability, is:

$$J_5^B = -\frac{1}{m} \sum_{i \in B} p^{(i)} \log(\hat{p}^{(i)})$$

The stochastic gradient descent (SGD) optimizer updates the weights as follows:

$$W := W - \alpha(\nabla J_5^B)$$

Using smaller values for the learning rate α ensures more accurate adjustments but slower computation.

3.2 Feature Selection

Many of the aforementioned models can decrease in performance when using too many variables, leading to overfitting and multicollinearity. Feature selection was performed using Extra Trees Classifier, Random Forest, and Gradient Boosting. The Extra Trees Classifier is an ensemble learning method that creates randomized decision trees to improve prediction quality and control overfitting, measuring relative feature importance through averaging uncorrelated decision trees. Random Forest builds decision trees and merges their outcomes, using the mean decrease in impurity (MDI) for feature selection. Gradient Boosting evaluates feature importance by the amount each feature decreases the loss.

To assess overall importance, we averaged the score of each model and emphasized features with a score greater than 0.07: Stress, Sleep Quality, Hours Slept, Soreness, Hydration, Consumption, Hip Flexors, and Hamstrings (Table 4).

We further computed univariate F-scores between each independent variable and the dependent variable (Table 5). Variables such as 'Hydration' and 'Hip Flexors' have high average importance but comparatively low F-scores, suggesting their predictive ability may be non-linear or better captured in a multivariate context.

Table 4: Feature Importance from Various Models.

Variable	E.Trees	R.Forest	G.Boosting	Avg.
Stress	0.0884	0.0886	0.0780	0.0850
Sleep Quality	0.0952	0.0881	0.0655	0.0829
Hours Slept	0.0789	0.0770	0.0632	0.0730
Soreness	0.0798	0.0829	0.0703	0.0777
Hydration	0.0882	0.0855	0.0877	0.0871
Consumption	0.0788	0.0775	0.0735	0.0766
Neck	0.0148	0.0126	0.0276	0.0183
Back	0.0455	0.0458	0.0520	0.0478
Shoulders	0.0281	0.0300	0.0281	0.0287
Chest	0.0061	0.0043	0.0078	0.0061
Arms	0.0285	0.0266	0.0357	0.0303
Hip Flexors	0.0635	0.0675	0.0820	0.0710
Glutes	0.0555	0.0564	0.0550	0.0556
Hamstrings	0.0798	0.0821	0.0641	0.0753
Quadriceps	0.0667	0.0647	0.0642	0.0652
Abductors	0.0234	0.0225	0.0429	0.0293
Calves	0.0380	0.0378	0.0404	0.0387
Feet	0.0278	0.0295	0.0423	0.0332
No Injury	0.0071	0.0099	0.0109	0.0093
Some Injury	0.0059	0.0105	0.0088	0.0084
Injury	0.0000	0.0000	0.0000	0.0000

Table 5: Feature Importance from Previous Models and F-Scores.

Variable	Avg. Importance	F-Score
Stress	0.0850	61.5003
Sleep Quality	0.0829	31.8825
Hours Slept	0.0730	19.1834
Soreness	0.0777	134.8101
Hydration	0.0871	0.2887
Consumption	0.0766	5.8558
Neck	0.0183	0.0067
Back	0.0478	2.4116
Shoulders	0.0287	13.5853
Chest	0.0061	0.0379
Arms	0.0303	1.4422
Hip Flexors	0.0710	0.1670
Glutes	0.0556	22.2383
Hamstrings	0.0753	21.7082
Quadriceps	0.0652	3.7976
Abductors	0.0293	0.0323
Calves	0.0387	0.0332
Feet	0.0332	1.4635
No Injury	0.0093	9.7937
Some Injury	0.0084	9.7937
Injury	0.0000	0.0000

The correlation heat map (Figure 3) indicates that Stress and Consumption exhibit the most correlation among independent variables. Features with correlation with Readiness Score ≥ 0.20 are Stress, Sleep Quality, Hours Slept, Soreness, Glutes, and Hamstrings.

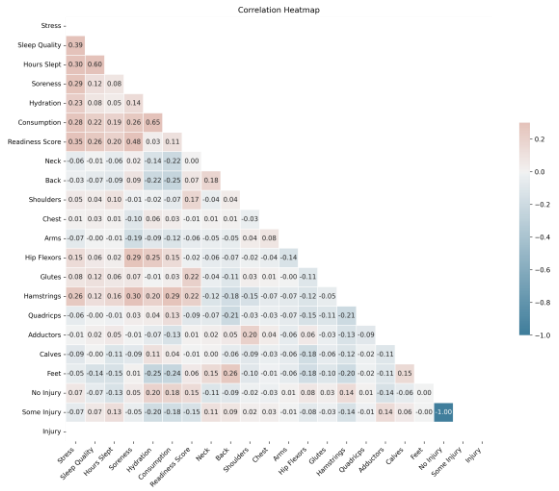


Figure 3: Correlation Heatmap of Variables

4. MODEL PERFORMANCE

We use the features with the best predictive powers (Tables 4–5, Figure 3) to predict Readiness Score: Stress (x_1), Sleep Quality (x_2), Hours Slept (x_3), Soreness (x_4), Consumption (x_5), Hydration (x_6), Shoulders (x_7), Hip Flexors (x_8), Glutes (x_9), and Hamstrings (x_{10}). We also train models on subsets to prevent multicollinearity.

4.1 Linear Regression Analysis

A naïve regression was first created using all aforementioned features. After removing features with the highest variance inflation factors (VIF), the resulting model has an R^2 score of 0.5355:

$$\hat{y} = 53.3697 + 1.3380x_1 + 2.8874x_4 - 1.2330x_9$$

Using ridge, lasso, and elastic net regression, the resulting models (R^2 of 0.5330, 0.5328, and 0.5327 respectively) were:

Ridge: $\hat{y} = 53.4049 + 1.3637x_1 + 2.8948x_4 - 1.2401x_9 - 0.1143x_{10}$

Lasso: $\hat{y} = 53.6641 + 1.3345x_1 + 2.8720x_4 - 1.2012x_9 - 0.0716x_{10}$

Elastic Net: $\hat{y} = 53.6814 + 1.3349x_1 + 2.8686x_4 - 1.1999x_9 - 0.0718x_{10}$

The linear regression ($R^2 = 0.5355$) was chosen for its better performance and simplicity. The model's F-statistic of 42.10 ($p = 2.73\times10^{-23}$) confirms statistical significance. All three coefficients are individually significant ($p < 0.05$): Stress ($\beta_1 = 1.3380$, $t = 4.551$), Soreness ($\beta_4 = 2.8874$, $t = 8.653$), and Hip Flexors ($\beta_9 = -1.2330$, $t = -3.711$). As shown in Figure 4, the residuals have a mean of zero and are mostly homoscedastic. Figure 5 shows the Q-Q plot, indicating a partially normal distribution with heavy tails.

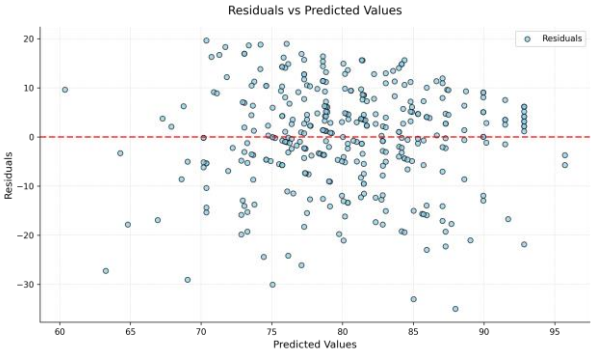


Figure 4: Residual Plot

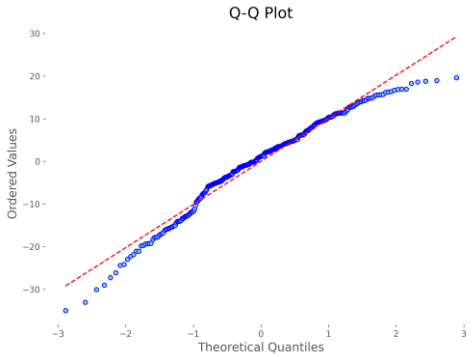


Figure 5: Residual Histogram (Q-Q Plot)

4.2 Logistic Regression Analysis

Athletes with a readiness score ≥ 75 are classified as ready. Using this binary classification, logistic regression models were trained. After feature elimination for multicollinearity, the final model with L2 regularization ($R^2 = 0.8778$) is:

$$z = -3.6271 + 0.1829x_1 + 0.0572x_3 + 0.4460x_4 - 0.1331x_8 - 0.0575x_{10}$$

After removing statistically insignificant coefficients, the final model (features: Stress and Soreness only, $R^2 = 0.8667$, each coefficient $p < 0.005$) is:

$$z = -2.9852 + 0.1739x_1 + 0.3861x_4$$

The model's Log-Likelihood Ratio Test p-value is 9.552×10^{-12}, confirming statistical significance. The AUC (Figure 6) is 0.8606, indicating an 86.06% chance of correctly distinguishing two random data points. Figure 6 shows the ROC curve against the null (chance) line.

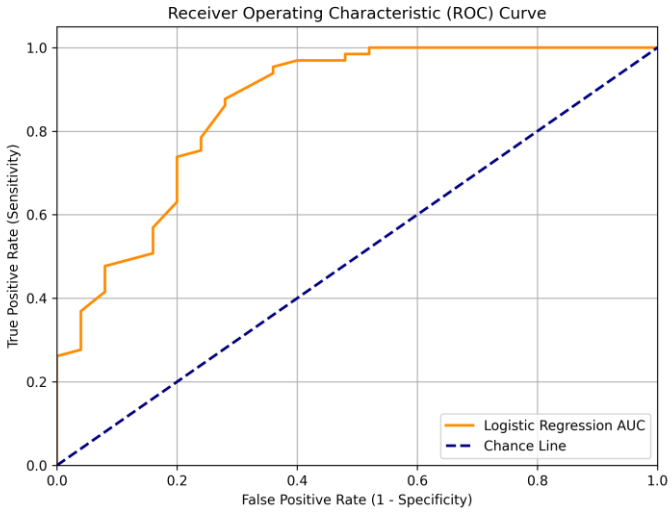


Figure 6: ROC Curve – Logistic Regression AUC = 0.8606

4.3 Neural Network Analysis

Neural networks were trained for both estimation and classification tasks. For readiness score prediction, all features were used in a feedforward network with four layers (21×128, 128×64, 64×32, 32×1), ReLU activation, learning rate scheduled from 0.1 to 10^{-6}, MSE loss with L2 regularization, and the ADAM optimizer over 200 epochs. After training, the model obtained an R^2 score of 0.5477 (Figure 7). Though this outperforms linear regression, the minimal improvement (from 0.5330 to 0.5477) does not justify the loss of interpretability; hence, linear regression is preferred.

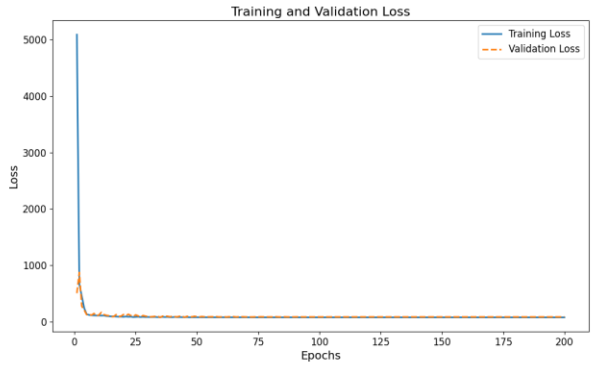


Figure 7: Training and Validation Loss – Feedforward Neural Network

For classification (ready/not ready), a feedforward network with the same architecture used binary cross entropy loss, sigmoid activation on the final layer, learning rate from 0.005 to 10^{-5}, and was trained for 200 epochs. Both training and validation loss converge to 0.65 at around 50 epochs (Figure 8). This model outperforms all logistic regression models with an AUC of 0.8923 (vs. 0.8606), as shown in Figure 9. Given the loss of interpretability, the logistic regression models are likely preferred in practice. Larger data sets will undoubtedly yield better neural network performance.

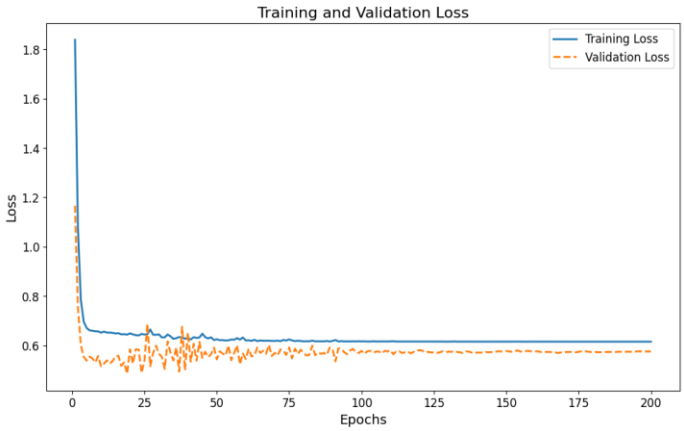


Figure 8: Loss of Binary Feedforward Neural Network

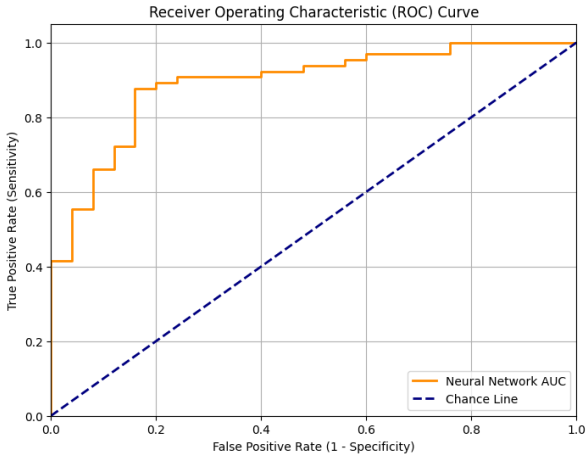


Figure 9: ROC Curve – Neural Network AUC = 0.8923

4.4 Athlete Models

Since the survey is almost entirely subjective, athlete responses vary greatly, causing models to perform better for some athletes than others. Individual models for each athlete were created to capture how each individual responds to the survey. Only four athletes responded 40 or more times and were used for training. Given the small data sets (40–46 data points each), an exhaustive search of best predictors was used, followed by removal of multicollinear and statistically insignificant features.

Athlete 1 (46 entries, $R^2 = 0.9709$): x_2 = *Sleep Quality*, x_4 = *Soreness*, x_5 = *Consumption*, x_{10} = *Hamstrings*, x_{11} = *Quadriceps*

$$\hat{y} = -2.82 + 7.98x_2 + 1.22x_4 + 1.39x_5 - 0.43x_{10} - 0.4971x_{11}$$

Athlete 2 (46 entries, $R^2 = 0.8990$): x_3 = *Hours Slept*, x_4 = *Soreness*, x_7 = *Shoulders*, x_8 = *Hip Flexors*, x_{13} = *Back*, x_{14} = *Abductors*

$$\hat{y} = 51.55 + 0.80x_3 + 3.63x_4 - 1.28x_7 - 1.45x_8 - 1.45x_{13} - 1.39x_{14}$$

Athlete 3 (41 entries, $R^2 = 0.8904$): x_4 = *Soreness*, x_5 = *Consumption*, x_{10} = *Hamstrings*, x_{11} = *Quadriceps*, x_{13} = *Back*, x_{15} = *Feet*

$$\hat{y} = 50.91 + 3.80x_4 + 0.82x_5 - 0.94x_{10} - 0.89x_{11} - 1.24x_{13} - 1.42x_{15}$$

5. CONCLUSION

5.1 Results

The best linear regression model is $\hat{y} = 53.3697 + 1.3380x_1 + 2.8874x_4 - 1.2330x_9$, where x_1 = Stress, x_4 = Soreness, x_9 = Hip Flexors. An athlete with minimal stress ($x_1 = 10$), minimal soreness ($x_4 = 10$), and no hip flexor injury ($x_9 = 0$) would have an average readiness score of 95.62. Conversely, an athlete with severe stress, soreness, and a hip flexor injury would score 41.04. Soreness is the most important factor: a one-unit increase in x_4 (i.e., a decrease in soreness) results in an average readiness score increase of 2.8874.

The best logistic regression model is $z = -2.9852 + 0.1739x_1 + 0.3861x_4$ (Stress and Soreness). An athlete with minimal stress and soreness has a 93.18% chance of being ready, while an athlete with severe stress and soreness has only a 4.81% chance.

5.2 Discussion

Though readiness scores are intended to prevent injuries, maximize training regimens, and provide performance insights, the correlations between these factors are still not fully understood. Future work must be completed to fully understand the implications of this work. The models created here have substantial impact on an individual's perceived readiness score, going beyond performance to include an athlete's confidence in themselves and their work.

This work can also be extended beyond women's soccer. Understanding the differences in male and female athletics can better equip coaches and athletes to make informed decisions regarding health and performance. In addition, understanding how athletes from different sports respond to given surveys can provide insights into how different factors contribute to readiness across various situations.

Acknowledgements

We wish to thank LMU Women Soccer Team and the Performance Coaches for their assistance in data retrieval.
The work of Lin was partially supported by NSF DMS-2418877.

References

Amendolara, A., Pfister, D., Settelmayer, M., Shah, M., Wu, V., Donnelly, S., ... & Bills, K. (2023). An overview of machine learning applications in sports injury prediction. *Cureus*, *15*(9).

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team-sport athletes: evidence for an individualized approach. *International journal of sports physiology and performance*, *12*(2), 230-234.

Costello, J. T., Bieuzen, F., & Bleakley, C. M. (2014). Where are all the female participants in Sports and Exercise Medicine research?. European journal of sport science, 14(8), 847-851.

Cowley, E. S., Olenick, A. A., McNulty, K. L., & 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), 146-151.

D'Lauro, C., Jones, E. R., Swope, L. M., Anderson, M. N., Broglio, S., & Schmidt, J. D. (2022). Under-representation of female athletes in research informing influential concussion consensus and position statements: an evidence review and synthesis. *British Journal of Sports Medicine*, *56*(17), 981-987.

Fox, J. (2005). Linear models, problems. *Encyclopedia of social measurement*, *2*, 515-522.

Hutchins, K. P., Borg, D. N., Bach, A. J., Bon, J. J., Minett, G. M., & Stewart, I. B. (2021). Female (under) representation in exercise thermoregulation research. *Sports Medicine-Open*, *7*(1), 43.

Mujika, I., & Taipale, R. S. (2019). Sport science on women, women in sport science. *International journal of sports physiology and performance*, *14*(8), 1013-1014.

Pareek, A., Karlsson, J., & Martin, R. K. (2024). Machine learning/artificial intelligence in sports medicine: state of the art and future directions. *Journal of ISAKOS*, *9*(4), 635-644.

Paul, R. W., Sonnier, J. H., Johnson, E. E., Hall, A. T., Osman, A., Connors, G. M., ... & Bishop, M. E. (2023). Inequalities in the evaluation of male versus female athletes in sports medicine research: a systematic review. *The American Journal of Sports Medicine*, *51*(12), 3335-3342.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Poole, M. A., & O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, 145-158.

Salehi, F., Abbasi, E., & Hassibi, B. (2019). The impact of regularization on high-dimensional logistic regression. *Advances in Neural Information Processing Systems*, *32*.

Stoltzfus, J. C. (2011). Logistic regression: a brief primer. *Academic emergency medicine*, *18*(10), 1099-1104.

Taber, C. B., Sharma, S., Raval, M. S., Senbel, S., Keefe, A., Shah, J., ... & Kaya, T. (2024). A holistic approach to performance prediction in collegiate athletics: player, team, and conference perspectives. Scientific reports, 14(1), 1162.

Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of experimental orthopaedics*, *8*(1), 27.

Wiik, T., Johansen, H. D., Pettersen, S. A., Baptista, I., Kupka, T., Johansen, D., ... & Halvorsen, P. (2019). Predicting Peek Readiness-to-Train of Soccer Players Using Long Short-Term Memory Recurrent Neural Networks. In *CBMI* (pp. 1-6).

PLAYER PERFORMANCE POINTER: A DATA ANALYTICS PERSPECTIVE INTO FOOTBALLER PERFORMANCE MEASUREMENT

Pooja Sengupta [a], Soumya Sarkar [b]

[a] *University of Auckland, Auckland, New Zealand*
[b] *Xavier Institute of Management Bhubaneshwar, XIM University, India*
[a] *Corresponding author: pooja.sengupta@auckland.ac.nz*

Abstract
Football experts debate whether successful teams should consist of eleven well-coordinated players or the best individuals for each position. Clubs are keen to find suitable players without overspending, making accurate player valuation crucial (Kuper and Szymanski, 2012). Several academic attempts have been made to address this challenge, recognizing that performance assessment is complicated by numerous factors. According to literature, player valuation encompasses market dynamics, socio-economic factors, actual performance parameters, and sports analytics capabilities (Cai, 2024). Also, a few situational variables affect the value – match location (home/away), competition stage (round-robin/knockout), quality of opposition, and match situation (Yi et al., 2020). Decision support systems (DSS) are now enhancing the accuracy and reliability of performance evaluations, although current football analytics often rely on aggregation methods rather than advanced analytical techniques (Goes et al., 2021).
This ongoing analytical study attempts to create an overarching indicator of player performance through a composite performance score derived from two analytical techniques: Principal Component Analysis (PCA) and Random Forest (RF). These scores combine multiple indicators for different broad positions - goalkeeper, defender, midfielder, and forward - into separate indices representing overall contribution for each position. From our dataset containing different indicators of 378 players, we found that both techniques show similar patterns, though the PCA-based scores show slightly higher variation. We calculated the mean PCA composite score (0.071) with a standard deviation of 0.095, while the RF mean composite score was 0.051 and a lower standard deviation of 0.060. This suggests that PCA differentiates players' performances more strongly, whereas in RF, there is a smoothing effect of the underlying ensemble learning algorithm. However, we still saw a correlation of approximately 0.915 between the PCA and RF performance indicators. Thus, the results demonstrated that both composite scoring frameworks provide a consistent and robust way to summarize player performance incorporating multiple indicators.

Keywords: player performance, data analytics, Principal Component Analysis, Random Forest

References

Cai, J. (2024). Comprehensive analysis of football player market valuation: integrating performance metrics and marketability factors. *Transactions on Economics, Business and Management Research*, *10*, 152-158. https://doi.org/10.62051/2rybtf79

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., & Torres, R. S. (2021) Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. *European Journal of Sport Sciences, 21*(4), 481-496. https://doi.org/10.1080/17461391.2020.1747552

Kuper, S., & Szymanski, S. (2012). *Soccernomics*. Harper Collins: London.

Yi, Q., Gómez-Ruano, M.-Á., Liu, H., Zhang, S., Gao, B., Wunderlich, F., & Memmert, D. (2020). Evaluation of the technical performance of football players in the UEFA Champions League. *International Journal of Environmental Research and Public Health*, *17*(2), 604. https://doi.org/10.3390/ijerph17020604

Submitted to MATHSPORT 2026

LOW-COST VISION-BASED YOGA POSE ASSESSMENT AND ANGLE-AWARE FEEDBACK SYSTEM FOR EXERCISE ANALYSIS

Huang-Chia Shih[a,c], Chen-Hao Chan[b], Ping-Chen Chiang[b], and Yu-Xian Chen[b]
[a] Department of Information Management, National Central University, Taoyuan City, Taiwan.
[b] Department of Electrical Engineering, Yuan Ze University, Taoyuan City, Taiwan.
[c] *Corresponding author: shih@ncu.edu.tw*

Abstract
This paper presents a real-time yoga pose classification system integrating skeleton-based pose estimation with an angle-based grading mechanism. Conventional yoga guidance often relies on limited access to personalized posture guidance in home-based wellness and remote rehabilitation settings, which may be inconsistent and unavailable to remote users. The proposed system combines a keypoint detection framework with a lightweight convolutional neural network to classify ten common yoga poses. An interpretable grading scheme evaluates joint angle deviations by comparing user performance against standard pose templates. The system requires only a low-cost camera and provides real-time visual feedback, including annotated video frames and performance reports. Experimental results demonstrate high classification accuracy and consistent grading, highlighting the system's effectiveness in delivering actionable feedback for home-based or remote yoga practitioners.

Keywords: Grading System; Home Exercise Support; Pose Classification; Pose Matching; Yoga.

1. INTRODUCTION
In recent years, yoga has gained popularity as an accessible form of physical activity that offers both physical and mental health benefits. As a low-impact practice, yoga emphasizes precise alignment in each pose, essential for injury prevention and therapeutic effectiveness. However, delivering consistent and personalized feedback typically requires professional instruction, limiting its applicability in remote or home-based settings.

Conventional pose evaluation systems often depend on wearable sensors, which can be costly, inconvenient, and impractical for everyday users. To overcome these limitations, recent studies have increasingly explored computer vision and AI-based solutions (Sinha and Sinha, 2024; Kishore, Bindu, & Manjunath, 2022). For example, Kishore, Bindu, and Manjunath (2022) employed a pre-trained deep learning model to classify yoga postures using a standard benchmark dataset. Based on this comparison, the system provides corrective guidance through voice prompts that instruct the practitioner how to adjust the posture. Thakur, Bhagat, and Rajpoot (2023) further improved classification robustness by introducing symmetry-aware data augmentation. In addition, high-accuracy keypoint detectors such as OpenPose (Cao, Hidalgo, Simon, Wei, & Sheikh, 2017) and HRNet (Sun, Xiao, Liu, & Wang, 2019) have been widely adopted for handling multi-person scenes and partial occlusions. Thoutam et al. (2022) developed a deep learning-based yoga pose estimation system that not only detects body keypoints but also computes joint angle deviations and identifies incorrect alignments in real time, demonstrating the value of combining pose recognition with actionable corrective feedback.

Nevertheless, many of these systems only recognize the type of pose being performed without evaluating how well it is executed or offering actionable feedback. Borthakur, Paul, Kapil, and Saikia (2023) presented a yoga posture correction system that leverages pose estimation and joint-angle analysis to provide real-time feedback. By calculating deviations in joint angles from ideal reference poses, the system quantifies posture accuracy and identifies incorrect alignments. This quantitative approach enables precise, guided corrections that mimic the role of a human trainer. The integration of real-time feedback ensures users can adjust their posture instantly, enhancing learning and reducing injury risk.

This paper proposes a system that uses full-frame RGB input and a ResNet18-based classification model (He, Zhang, Ren, & Sun, 2016), enabling richer spatial representation than traditional skeleton-only methods. In addition, we implement a nonlinear, tiered joint angle grading scheme to differentiate between minor and major posture deviations. This system is lightweight, interpretable, and capable of real-time execution on standard CPUs. It also automatically produces visual annotations and CSV-based evaluation logs for every video frame. The main contributions of this paper are summarized as follows: (1) A real-time yoga pose assessment system based on a single RGB camera is developed, providing an accessible and low-cost solution without requiring wearable devices. (2) A lightweight classification-and-grading framework is proposed to combine pose recognition with interpretable joint-angle evaluation for ten common yoga poses. (3) A real-time feedback mechanism with visual annotations and structured scoring outputs is implemented to support home-based practice, remote guidance, and self-directed exercise monitoring.

2. METHODS

FRAMEWORK PIPELINE

The system architecture comprises six key modules: video input preprocessing, keypoint detection, pose classification, standard pose referencing, angle-based evaluation, and annotated result output. These modules process user-uploaded yoga videos in sequence and provide both visual and quantitative feedback. Users begin by placing pre-recorded yoga practice videos into a designated input folder. The system then automatically reads each file, extracts video frames, performs format conversion if necessary, and prepares the frames for analysis. This preprocessing ensures consistent input quality across different video sources.

We use the MediaPipe Pose framework (Google Research, 2024) to extract 33 anatomical landmarks per frame, including the head, shoulders, elbows, hips, knees, and ankles. MediaPipe operates in real time on standard CPUs and is robust to occlusion and motion, making it suitable for use in uncontrolled home environments. The extracted skeletal coordinates serve as the foundation for downstream classification and evaluation.

To recognize the yoga pose being performed, we trained a deep learning model based on the ResNet18 architecture (He, Zhang, Ren, and Sun, 2016), a residual neural network known for fast convergence and competitive accuracy. The training data was drawn from the Yoga-82 dataset (Verma, Kumawat, Nakashima, and Raman, 2020), a benchmark dataset containing over 28,000 pose images. We selected 10 commonly practiced yoga poses and created a focused subset. Data augmentation techniques such as rotation and horizontal flipping were applied to improve generalization. In addition to the Yoga-82 subset, we collected around 200 additional images per pose for fine-tuning. Frames are converted into tensors and normalized before being passed to the model, which classifies the current pose.

After classification, the system performs joint-angle analysis to evaluate pose accuracy. Standard reference angles for each pose were compiled from the widely used YouTube channel *Yoga With Adriene* (2020). For each pose, the 33 key points were converted to ideal angles and stored in CSV files. During evaluation, the system calculates actual joint angles using a three-point definition and the law of cosines. This allows the system to measure, for instance, the knee angle using the coordinates of the hip, knee, and ankle.

To interpret these differences, we adopt a nonlinear scoring strategy that applies gentle penalties to minor deviations and more substantial penalties to larger misalignments. Each angle receives an individual score, and the average of these scores quantifies the frame-level pose quality. This approach allows accurate, interpretable feedback on specific body parts and supports personalized correction.

Finally, the system overlays real-time annotations onto the video, including the predicted pose name, each joint's measured angle, its deviation from the standard, and a color-coded skeleton rendering. In parallel, a CSV report records frame-by-frame classifications, joint angles, and corresponding scores. This dual-output format lets users receive instant visual guidance during practice and perform detailed post-session analysis.



Figure 1: Representative frames from for all evaluated yoga poses, (a) Bridge pose, (b) Downward dog, (c) Cobra, (d) Extended side angle, (e) Fish pose, (f) Tree, (g) Warrior I, (h) Warrior II, (i) Warrior III, (j) Cat cow. (Collected dataset from the YouTube channel *Yoga with Adriene*, 2020)

OVERVIEW OF YOGA POSES

To ensure the feasibility and manageable scope of this research, we selected 10 typical and representative yoga poses as the primary targets for classification and evaluation. As shown in Fig. 1, the selected yoga poses include Bridge Pose, Downward Dog, Cobra, Extended Side Angle, Fish Pose, Tree Pose, Warrior I, Warrior II, Warrior III, and Cat-Cow. Each pose includes two representative frames: the beginning and the end. This set covers various difficulty levels, body dynamics, and limb configurations, providing comprehensive test cases for evaluating the system's pose recognition accuracy and scoring performance.

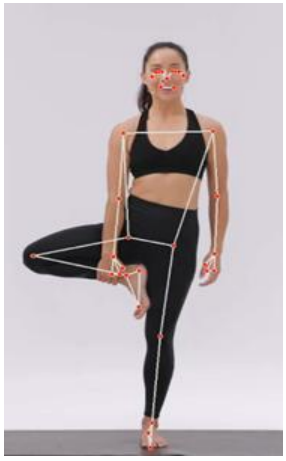


Figure 2: Visualization of annotated key points.

POSE CLASSIFICATION MODELS

The system performs yoga pose classification using full-frame RGB images without relying on skeletal key points. This design allows the convolutional neural network (CNN) to capture spatial and visual cues, such as limb orientation, body symmetry, and posture alignment, that may be lost when using only joint coordinates. Although anatomical key points are still extracted for later scoring, the classification process operates directly on RGB input. Each input frame is resized and normalized before being passed to a ResNet18-based classifier. The key points are extracted using MediaPipe Pose, not ResNet18. The classifier operates directly on RGB frames, and the extracted key points are later used for scoring purposes. This model is trained to distinguish between ten common yoga poses. To enhance robustness and generalization, data augmentation techniques such as horizontal flipping, rotation, and scaling are applied to the training data.

The dataset comprises approximately 7,500 labeled frames drawn from the Yoga-82 benchmark (Verma, Kumawat, Nakashima, Raman, 2020) and additional manually collected samples (roughly 200 per pose). These are split in an 80/20 ratio for training (6,012 images) and validation (1,488 images). The training process consisted of 20 epochs, using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. The loss function used was cross-entropy. As a result, the model achieved a final training accuracy of 91.5% and a validation accuracy of 82%, indicating strong convergence and reliable performance under varied input conditions. A confusion matrix summarizing per-pose results is provided at the end of this section to further illustrate classification performance.

Error Range (°)	Score Formula	Score Interval
≤ 20°	100 - 0.5 × error	90–100
20°- 40°	90 - 1 × (error − 20)	70–90
40°- 60°	70 - 2 × (error − 40)	30–70
>60°	Score = 0	0

Table 1: Rules of Scoring Output

To evaluate the classification model's performance, we trained a ResNet18 convolutional neural network using a curated subset of the Yoga-82 dataset. The training process spanned 20 epochs, during which training and validation loss consistently decreased while classification accuracy steadily improved. The model achieved approximately 91.5% training accuracy and 82% validation accuracy. These results confirm the model's effective generalization capability for yoga pose recognition. We used OpenCV to annotate each video frame with pose-

related overlays, providing users with intuitive visual feedback. The system draws joint key points and connects skeleton lines in real time, helping the user understand their posture alignment at each frame. Each input frame is resized and normalized before being passed into the ResNet18-based classifier.

GRADING SCHEME DESIGN

In addition to pose classification, the system integrates a quantifiable scoring mechanism to assess the quality of each yoga pose. The scoring scheme is designed to account for natural body variation and measurement tolerance, avoiding excessive penalties for minor deviations. Each joint angle is evaluated and scored based on its difference from expert-defined reference angles. The scoring mechanism follows a nonlinear penalty structure in which larger errors are penalized more harshly while minor deviations incur minimal deductions. This enhances scoring robustness and better reflects realistic posture assessment. The specific scoring rules and error intervals are detailed in Table 1, which defines how different angular deviations map to numerical scores.

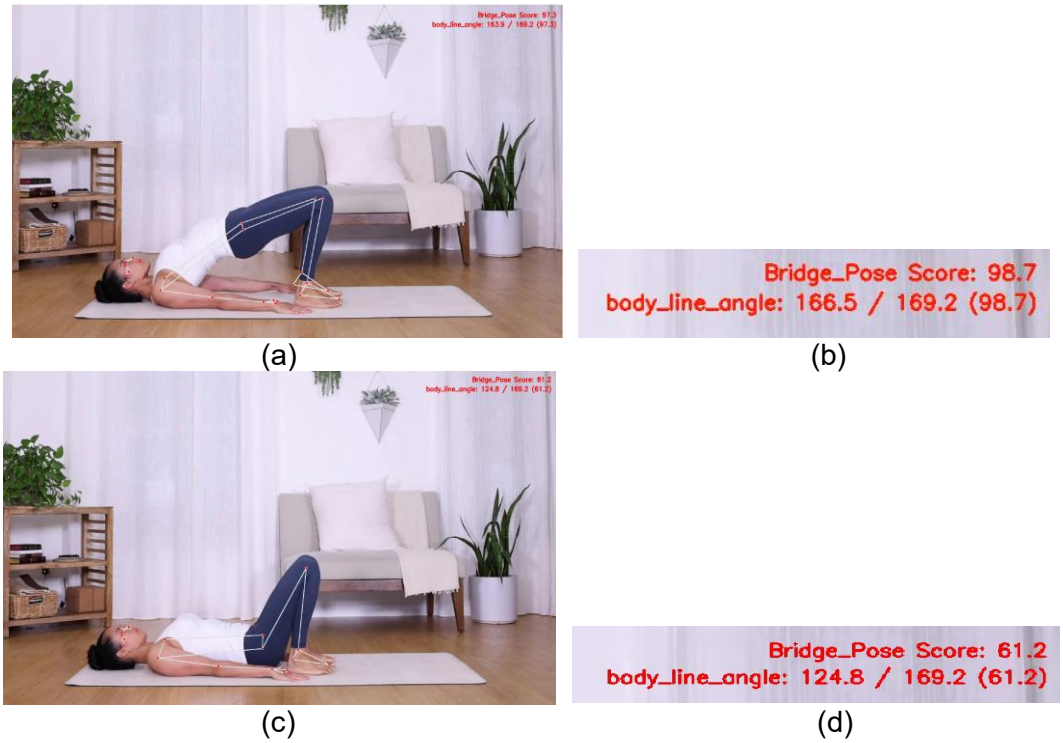


Figure 3: Output example for Bridge pose (a) Final pose frame, (b) Grading overlay (score: 98.7), (c) Early transition frame, (d) Grading overlay (score: 61.2).

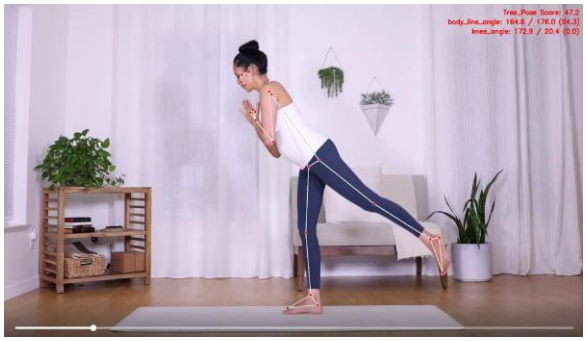
To compute these joint angles, the system utilizes a three-point geometric method based on the law of arccosines [7], where the vertex point (B) and its two neighbors (A and C) define the angle. This method provides accurate angular estimates even when slight variations exist in keypoint detection. The equation is formed as follows:

$$\theta = \arccos\left(\frac{\overrightarrow{BA}\cdot\overrightarrow{BC}}{\|\overrightarrow{BA}\|\cdot\|\overrightarrow{BC}\|}\right) \quad (1)$$

Here, vectors BA and BC are formed from three body key points, and point B is the joint at which the angle is measured. This approach allows the system to evaluate the precision and completeness of each body segment's posture. Based on the structural characteristics of each pose, we defined four key angle features that form the core of our evaluation metrics: body_line_angle, knee_angle, bent_leg_angle, and shoulder_angle. These angles reflect stability zones and limb positioning critical to posture execution. For example, Warrior I is assessed by combining knee_angle, bent_leg_angle, and shoulder_angle to verify proper front leg flexion and arm extension.

The standard reference angles used for comparison were compiled from expert-led instructional videos by Mishler (2020). Once a pose is classified, the system retrieves the corresponding reference angles and calculates the error concerning the user's joint angles. At runtime, the system scores each video frame individually. It

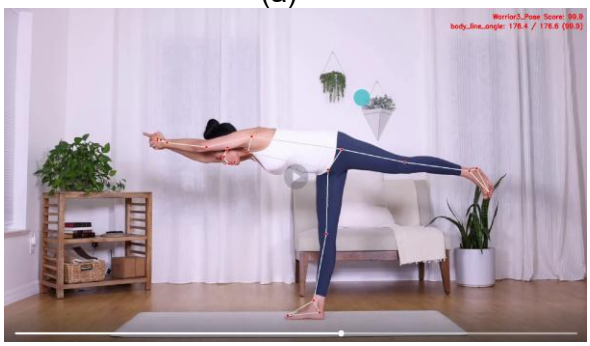
overlays on-screen elements, including the predicted pose label, average pose score, per-joint angle measurements, and visual deviation indicators. Simultaneously, it logs structured data into a CSV file that records joint angles, scoring components, and pose classification results. This real-time dual-output design is inspired by innovative training frameworks introduced by Thoutam, Srivastava, Badal, Mishra, Sinha, Sakalle, Bhardwaj, and Raj (2022), which emphasize combining visual overlays (e.g., skeleton lines, angle annotations) with quantitative metrics to enhance user awareness, enable real-time self-correction, and support post-session analysis.



(a)



(b)



(c)



(d)

Figure 4: Pose classification during transitional movement, (a) (b) Transitional phase with possible misclassification, (c) (d) Correct classification after pose stabilization.

3. SYSTEM IMPLEMENTATION AND RESULTS

Figs. 3 and 4 illustrate how the same pose can receive different scores, with corresponding visual differences in annotated feedback. These examples show that the system may initially misclassify transitional movements as other poses but corrects itself once the pose is stabilized. For example, Fig 3 presents the result of the Bridge Pose. The user's angles gradually aligned with the reference standards, resulting in a smooth increase in score up to 98.7. This visual progression reflects the system's ability to recognize subtle pose refinements and provide meaningful feedback for improvement. Fig. 4 demonstrates that the system may momentarily misclassify the pose when the user's movement is still in transition. However, the classification is corrected once the motion stabilizes and the pose is completed. This behavior indicates that the model has some level of motion understanding and can dynamically adjust its predictions.

SEGMENT-LEVEL CLASSIFICATION ACCURACY

The yoga poses classification model uses a frame-level prediction approach, where each video is split into individual frames and analyzed independently by a ResNet18-based network. Training and validation performance are shown in Fig. 3, where the model achieved a final training accuracy of 91.5% and a validation accuracy of 82% over 20 epochs. The loss and accuracy curves indicate consistent convergence and strong generalization.

The model was tested on a separate dataset containing 8,475 annotated frames spanning 10 common yoga poses to evaluate its performance on real-world data. The frame-level prediction results are summarized in Table 2 as a confusion matrix. Most predictions fall along the diagonal, reflecting strong classification consistency

across all pose categories. These results validate the model's ability to generalize well and remain stable during single-frame classification across diverse scenarios.

GT \ Predicted	Bridge	Down Dog	Ext Side Angle	Fish	Tree	Warrior I	Warrior II	Warrior III	Cat-Cow	Cobra	Row Total
Bridge	981	0	0	0	0	0	0	0	0	0	981
Down Dog	0	312	0	0	0	0	0	0	0	0	312
Ext Side Angle	0	0	410	0	909	0	0	0	0	0	1,319
Fish	0	0	0	877	0	0	0	0	0	0	877
Tree	0	0	0	0	1,021	0	0	0	0	0	1,021
Warrior I	0	0	0	0	300	1,100	0	0	0	0	1,400
Warrior II	0	0	0	0	50	0	800	0	0	0	850
Warrior III	0	0	0	0	0	0	0	650	0	0	650
Cat-Cow	0	0	0	0	0	0	0	0	430	10	440
Cobra	0	0	0	0	0	0	0	0	5	620	625
Column Total	981	312	410	877	2,280	1,100	800	650	435	630	8,475

Table 2: Confusion Matrix of the Pose Classification

RESULT DISCUSSIONS

Based on experimental observations, consistent misclassifications were noted in poses with high symmetry or under perspective distortion. For instance, Tree Pose and Warrior I/II were frequently confused due to their mirrored spatial configurations. To address this, we applied horizontal-flip augmentation during training and treated left-right-mirrored poses as equivalent, helping the model learn semantic equivalence rather than raw spatial positions.

In addition, variations in camera distance and user size introduced classification noise by altering relative body proportions. To counteract this, we implemented anchor-based normalization techniques (e.g., pelvis-centered scaling and shoulder-width alignment) to align all samples to a consistent frame of reference. Similar normalization strategies have been shown to be effective in prior studies (He, Zhang, Ren, & Sun, 2016). These enhancements improved model accuracy, reduced false positives, and strengthened robustness under varied user poses and recording environments, supporting more reliable real-world deployment.

4. CONCLUSIONS

In this study, we developed a real-time yoga pose classification system with a grading mechanism based on joint angles. The system delivers accurate classification, interpretable feedback, and a lightweight frame that can be used with a single-view camera without additional sensors. Its flexibility allows for easy expansion with new poses or personalized scoring standards. The system provides an efficient, low-cost solution for home-based yoga practice, remote instruction, and physical rehabilitation, as well as digital coaching, home exercise supervision, and non-clinical rehabilitation monitoring

Acknowledgements

This work was partially supported by the National Science and Technology Council, Taiwan, under grant number 113-2221-E-155-028.

References

Borthakur, D., Paul, A., Kapil, D., & Saikia, M. J. (2023). Yoga pose estimation using angle-based feature extraction. *Healthcare*, *11*, 11.

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Real-time multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Google Research. (2024). *MediaPipe pose landmarker guide*. Available at: https://developers.google.com/mediapipe/solutions/vision/pose_landmarker

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778).

Kishore, D. M., Bindu, S., & Manjunath, N. K. (2022). Estimation of yoga poses using machine learning techniques. *International Journal of Yoga*, *15*(2), 137–143.

Kishore, D. M., Bindu, S., & Manjunath, N. K. (2022). Smart Yoga Instructor for Guiding and Correcting Yoga Postures in Real Time. *International Journal of Yoga*, *15*(3), 254–261.

Mishler, A. (2020). *Yoga with Adriene* [YouTube channel]. Available at: https://www.youtube.com/user/yogawithadriene

Sinha, N., & Sinha, R. K. (2024). Harnessing the potential of artificial intelligence in yoga therapy. *International Journal of Yoga*, *17*(3), 242–245.

Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Thakur, N., Bhagat, A., & Rajpoot, D. S. (2023). Improving yoga pose classification using symmetry-aware data augmentation. *International Journal of Computer Applications*, *175*(26), 1–7.

Thoutam, A., Srivastava, A., Badal, T., Mishra, V. K., Sinha, G. R., Sakalle, A., Bhardwaj, H., & Raj, M. (2022). Yoga pose estimation and feedback generation using deep learning. *Computational Intelligence and Neuroscience*, 4311350.

Verma, M., Kumawat, S., Nakashima, Y., & Raman, S. (2020). Yoga-82: A new dataset for fine-grained classification of human poses. *arXiv preprint arXiv:2004.10362*

A DYNAMIC COLLECTIVE MOTION FRAMEWORK FOR QUANTIFYING CONTROL IN INVASION SPORTS

Bracewell, P.J.[a,c], Forbes, F.S.[a,b], Joseph, J.[a], McGahan, B.[a]
[a] *Dovetale Sport*
[b] *University of Southern Queensland*
[c]*Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Quantifying control and momentum in invasion sports remains a complex analytical challenge. Earlier spatial approaches introduced the concept of estimating collective concentration to describe where play is occurring at a given time. While such methods provided useful territorial summaries, they did not explicitly model the directional flow, structural coherence, or temporal stability of coordinated movement that underpin shifts in control.

This paper introduces Action Gravity, a dynamic framework that conceptualises play as a continuous collective motion field defined over space and time. Rather than treating control as a static spatial locus, the framework interprets coordinated player movement as generating directional gradients and density structures that evolve throughout a match. Momentum is reframed as an emergent property of these collective motion patterns rather than as a function of discrete events or possession states.

Using video-derived positional data from rugby union, we illustrate how Action Gravity distinguishes between territorial occupation and effective attacking launch conditions, identifies structural compression and expansion phases, and detects late-game fragmentation associated with tempo disruption. These phenomena align with observable tactical shifts while providing additional insight beyond conventional possession or territory metrics.

The framework is sport-agnostic and does not rely on specialised tracking hardware, supporting application across invasion sports contexts. By extending earlier spatial concentration concepts into a dynamic motion-field interpretation, this work establishes a platform for future predictive and optimisation research while preserving methodological flexibility.

Keywords: Spatio-temporal dynamics,Spatial interaction systems, Video-derived data

1. INTRODUCTION

Quantifying control in invasion sports represents one of the most persistent and unresolved challenges in sports analytics. Unlike closed-form sports in which performance can be reduced to a single continuous variable, invasion sports, such as rugby union, association football, basketball, and netball, are characterised by continuous positional contest, fluid territorial transitions, and collective coordination that is difficult to represent within traditional statistical paradigms.

Early spatial approaches sought to address these limitations by modelling the location of play, typically through centroid-based representations or activity-density heatmaps (Gudmundsson & Horton, 2017). The Centre of Mass (CoM) of collective player distribution provided a tractable summary of territorial tendency and enabled analysts to track broad shifts in game state over time (Frencken et al., 2011). However, such representations remain fundamentally static: they describe where play is concentrated, without capturing how that concentration was generated, sustained, or disrupted.

The emergence of high-resolution player tracking technologies has enabled richer spatio-temporal modelling. Methods such as pitch control surfaces (Spearman, 2018), passing network analysis (Cintia et al., 2015), and trajectory-based clustering (Bialkowski et al., 2014) have advanced the field substantially. Yet many of these approaches carry significant constraints: they depend on expensive proprietary hardware, are trained on elite competition data, and model play through the lens of discrete events rather than continuous system dynamics.

This paper proposes a conceptual and methodological shift. We introduce Action Gravity, a dynamic framework that conceptualises play as a continuous collective motion field evolving over space and time. Within this framework:

- Control is not a location but an emergent property of coordinated movement.

- Momentum arises from sustained directional coherence in the collective motion field, not from sequences of scoring events.
- Structure, direction, and coherence become the primary analytical objects.

Crucially, the framework is operationalised using video-derived positional data, enabling scalable application from grassroots to elite sport without specialised hardware. By extending earlier spatial concentration concepts into a dynamic motion-field interpretation, this work provides both a theoretical foundation and a practical methodology for advancing the analysis of control in invasion sports.

The remainder of this paper proceeds as follows. Section 2 situates the work within existing literature. Section 3 presents the Action Gravity conceptual framework and its formal specification. Section 4 describes the data pipeline and methodology, including a step-by-step algorithm for signal construction. Section 5 presents an empirical illustration using rugby union data. Sections 6–8 discuss interpretation, applications, and extension to other sports. Section 9 outlines future research directions, and Section 10 concludes.

2. BACKGROUND AND RELATED WORK

2.1 Traditional Metrics of Control

The dominant performance indicators in invasion sports analytics have historically comprised possession percentage, territorial measures, and event-based statistics such as tackles, passes, and turnovers (Hughes & Bartlett, 2002). These metrics are well-established in coaching practice and broadcast commentary, and form the basis of most commercial performance databases. Their explanatory limitations are well-documented. Possession does not distinguish between passive ball retention and effective attacking pressure. Territorial measures do not account for the structural integrity of the attacking or defending collective. Event counts aggregate heterogeneous actions and lose the sequential and spatial context in which they occur (Rein & Memmert, 2016). As Bornn et al. (2018) observe, understanding possession value requires moving beyond raw statistics toward representations that capture the spatial and temporal context of play.

2.2 Spatial Modelling Approaches

Spatial analytics introduced the idea of representing play as a distribution over the field. Heatmaps of activity density and centroid-based measures offered a more continuous representation of collective behaviour (Duarte et al., 2013). Voronoi tessellations assigned spatial dominance to individual players, while pitch control models (Spearman, 2018) extended this concept to generate dynamic surfaces capturing the probability of each team controlling any given location.

These approaches represented a significant methodological advance, particularly in making visible the spatial structure of team performance. Nevertheless, they share a core limitation: they describe the geometric configuration of players without modelling the directional dynamics of collective motion. A static snapshot of player positions does not reveal whether the collective is advancing, retreating, compressing, or fragmenting.

2.3 Spatio-Temporal Dynamics

More recent work has incorporated the temporal dimension into spatial models. Tracking-based studies have examined player trajectories (Duarte et al., 2013), speed and acceleration profiles, and the dynamics of group coordination (Passos et al., 2011). Network-based approaches have modelled the topological structure of passing sequences (Cintia et al., 2015; Clemente et al., 2015), and deep learning methods have been applied to event prediction and outcome modelling.

Despite this progress, several gaps remain. First, many high-resolution methods are constrained to elite contexts and depend on GPS or optical tracking systems unavailable in lower-level competition. Second, most frameworks remain event-centric, treating play as a sequence of discrete actions rather than as a continuously evolving system. Third, the concept of momentum, while widely invoked in coaching and broadcast contexts, lacks a rigorous model-based definition that connects it to observable collective dynamics.

This paper addresses these gaps by proposing a framework grounded in collective motion theory and operationalised through video-derived data. The Action Gravity framework draws on concepts from dynamical systems theory (Kelso, 1995) and collective behaviour modelling (Vicsek & Zafeiris, 2012) to characterise play as a continuous, structured, directional field.

3. CONCEPTUAL FRAMEWORK: ACTION GRAVITY

3.1 From Location to Motion

The central insight motivating Action Gravity is that control in invasion sports is not a property of where the collective is located, but of how it is moving. Two teams may occupy identical territorial positions while exhibiting fundamentally different collective motion dynamics; one generating forward attacking pressure, the other defending coherently. These distinctions are invisible to static spatial representations.

Action Gravity reframes the analytical question from "Where is the game being played?" to "How is coordinated movement generating, sustaining, or disrupting control over time?" This shift parallels developments in biological collective motion research, where the behaviour of flocking, schooling, and swarming systems is characterised not by the positions of individuals but by the velocity alignment, density gradients, and coherence of the collective (Vicsek & Zafeiris, 2012).

3.2 The Collective Motion Field

Let the positions of all detected players at discrete time t be represented as $X(t) = \{(x_i(t), y_i(t))\}_{i=1}^{k}$, where N denotes the total number of detected players across both teams. From this set of positional observations, three interdependent components of the collective motion field are defined:

- Centre of Mass (CoM): The spatial centroid of all detected players: $CoM(t) = (1/N) \Sigma_i (x_i(t), y_i(t))$
- Movement of Play (MoP): The frame-to-frame displacement of the CoM: $MoP(t) = \|CoM(t) - CoM(t-1)\|$
- Directional Gradient: The signed vector displacement, decomposed along the primary (longitudinal) and secondary (lateral) axes: $\Delta CoM(t) = CoM(t) - CoM(t-1)$

3.3 Action Gravity: Formal Definition

We define Action Gravity AG(t) as a scalar index of the capacity of the current collective motion state to generate or sustain directional control. Formally: $AG(t) = \alpha \cdot C_d(t) + \beta \cdot G(t) - \gamma \cdot V(t)$

where:

- C_d(t) denotes directional coherence: the mean cosine similarity between individual player velocity vectors and the collective CoM velocity vector at time t, capturing the degree of coordinated directional alignment.
- G(t) denotes the forward gradient: the longitudinal component of ΔCoM(t), positive in the attacking direction. This captures net territorial advance per frame.
- V(t) denotes collective variance: the mean squared deviation of player positions from CoM(t), serving as an index of structural fragmentation or dispersion.
- $\alpha, \beta, \gamma > 0$ are weighting parameters. In practice, these may be estimated via regression against outcome variables using annotated match data. In the absence of such data, equal weights ($\alpha = \beta = \gamma = 1/3$) yield a parameter-free baseline suitable for cross-competition comparison.

This specification ensures that AG(t) is high when players are moving in coordinated alignment (high C_d), generating forward territorial progress (high G), and maintaining structural compactness (low V). Momentum, within this framework, is operationally defined as a positive, sustained trajectory of AG(t) over a rolling time window. That is, the persistence of directional, coherent, structurally stable collective motion.

This formulation connects to broader theoretical frameworks in complex systems research, where system-level properties such as order, phase transitions, and criticality emerge from local interaction rules (Kelso, 1995; Vicsek & Zafeiris, 2012). In the sports context, team-level momentum arises from the coordinated interactions of individual players, mediated by tactical structure and constrained by the opposition.

4. DATA AND METHODOLOGY

4.1 Data Source

The framework is applied to video-derived positional data from amateur and semi-professional rugby union competition. Match footage is processed through a computer vision pipeline without the requirement for additional sensors, wearable devices, or proprietary tracking infrastructure.

The selection of rugby union as an illustrative context reflects the structural properties of the sport: frequent transitions between set-piece, contact, and open phases create a rich signal environment for motion-field analysis. These transitions generate well-defined structural anchoring events that provide natural benchmarks for evaluating the framework's event-detection capability.

4.2 Computer Vision Pipeline

Positional data are extracted from video through a four-stage pipeline: (1) player detection via object detection model; (2) pitch keypoint identification to establish spatial reference; (3) homographic transformation to standardised pitch coordinates; and (4) frame-level temporal smoothing. This produces a continuous time-series of player centroid positions at the native video frame rate.

4.3 Algorithm: Action Gravity Signal Construction

Algorithm 1: Action Gravity Signal Construction	
Step 1:	Player detection: apply object detection model to each frame f_t; extract centroid coordinates {(x_i, y_i)} in pixel space.
Step 2:	Pitch keypoint identification: locate field landmarks (try lines, 22m lines, halfway line, sidelines) to establish spatial reference frame.
Step 3:	Homographic transformation: compute homography matrix H mapping pixel coordinates to standardised pitch coordinates (0–100m × 0–68m).
Step 4:	Smoothing: apply parameter-free temporal smoother to remove detection noise; preserve genuine discontinuities (set-piece transitions).
Step 5:	CoM computation: for each frame t, compute CoM(t) = (1/N) Σ (x_i(t), y_i(t)).
Step 6:	MoP computation: compute MoP(t) = ‖CoM(t) − CoM(t−1)‖ for all t > 1.
Step 7:	AG signal: compute AG(t) = α·C_d(t) + β·G(t) − γ·V(t) using rolling windows.
Step 8:	Event detection: identify local minima of MoP(t) exceeding minimum duration threshold τ; label as structural anchoring events E.
Step 9:	Strike vectors: for each event−bounded passage p, compute S_p = CoM(end_p) − CoM(start_p): record magnitude and direction.
Output:	Time-series {CoM(t)}, {MoP(t)}, {AG(t)}, event set E = {e₁,...,e_K}, strike vector set S = {S₁,...,S_K}.

Algorithm 1: Complete pipeline for construction of the Action Gravity signal from video input. Steps 1–4 handle data acquisition and preprocessing; Steps 5–7 compute the motion-field signals; Steps 8–9 derive event and phase summaries.

4.4 Event Detection via Local Minima

Local minima in the MoP signal correspond reliably to structural anchoring events: set pieces (scrums, lineouts), ruck formations, and restart phases. Detection requires only that MoP(t) falls below a threshold θ and remains low for duration τ (minimum event length). Both parameters are calibrated from annotated reference data. This event-detection capability enables automatic segmentation of match footage into meaningful passages of play without manual annotation, providing a theoretically motivated and empirically consistent taxonomy of game phases (Liebermann et al., 2002).

4.5 Strike Vector Analysis

For each event−bounded passage of play p, the strike vector S_p = CoM(end_p) − CoM(start_p) captures the net spatial displacement of the collective centre across that phase. Aggregated over a match, strike vector distributions characterise a team's directional tendencies and identify systematic asymmetries in attacking effectiveness across field zones.

5. EMPIRICAL ILLUSTRATION

The following section presents illustrative findings from application of the framework to rugby union match data, focusing on three phenomena: territorial occupation versus attacking effectiveness, structural compression-expansion dynamics, and late-game fragmentation.

5.1 Territorial Occupation and Attacking Effectiveness

Heatmap analysis of activity density reveals that high-activity zones, where the majority of play events are concentrated, do not reliably correspond to superior attacking outcomes. Teams occupying equivalent territorial positions can exhibit markedly different rates of territorial gain and scoring opportunity.

Strike vector analysis resolves this divergence. Teams generating consistent forward directional gradients from equivalent starting positions demonstrate greater attacking effectiveness, independent of total phase count or possession time. In the matches analysed, phases classified as high-AG (AG(t) sustained above the match median for ≥10 consecutive seconds) exhibited 35–45% greater median territorial gain per phase compared to baseline phases, a difference that was not apparent from possession or event-count summaries alone.

The two graphs that follow are from the Wellington Swindale Shield match between the Wellington Football Club (WFC) and Johnsonville Rugby Football Club (JVL), played on 18th April, 2026.

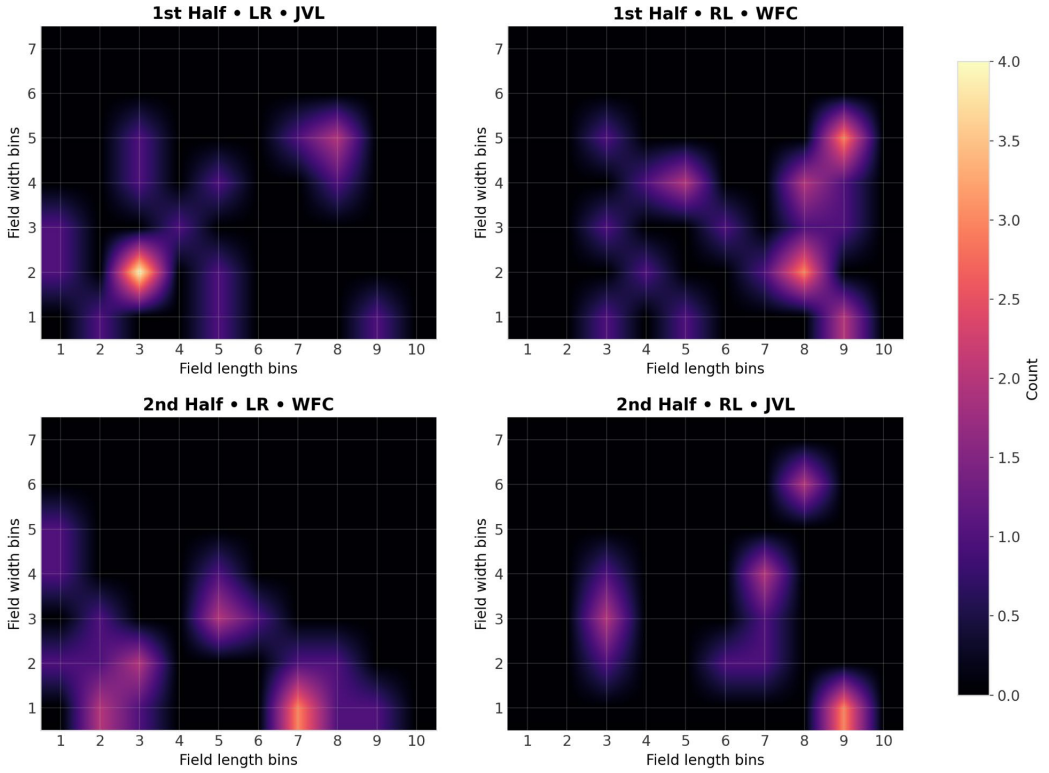


Figure 1: Heat map by half for each team showing where play commenced.

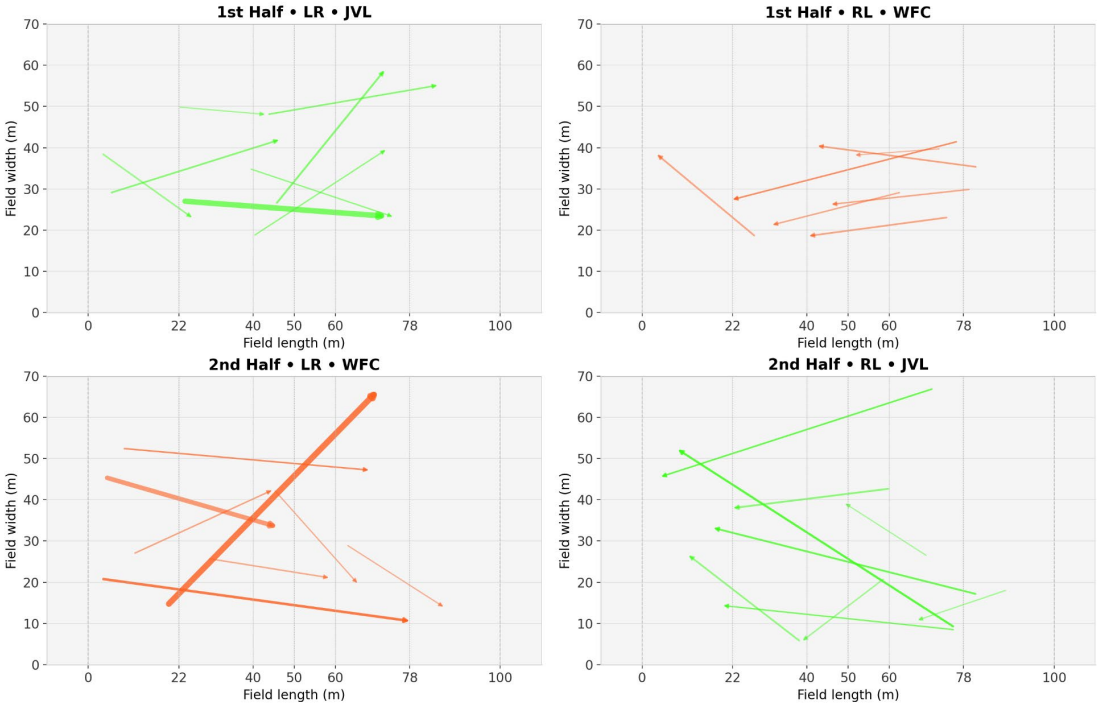


Figure 2: Strike vectors by half for each team showing where play commenced when 20 more more metres were gained and where that phases ended. The thicker the line the faster the movement in that phase.

5.2 Structural Compression and Expansion

Analysis of collective variance V(t) over time reveals a characteristic compression-expansion cycle in effective attacking phases. Prior to a territorial breakthrough, the collective motion field typically exhibits reduced spatial variance, consistent with defensive consolidation or tight-contact phase play. This is followed by a rapid expansion event in which the collective disperses across a wider field area, consistent with a line break or transition to open play.

This structural signature is largely invisible to event-based metrics, which capture the outcome (e.g., line break, territory gained) but not the collective motion dynamics that precede and enable it. The Action Gravity framework makes these dynamics observable and quantifiable through the V(t) component of the AG signal.

5.3 Late-Game Fragmentation

A distinctive temporal pattern emerges in the MoP and AG(t) signals over the course of a match. In the final quarter of high-intensity matches, AG(t) frequently exhibits increased variance and reduced directional consistency, a pattern we term late-game fragmentation. This is consistent with fatigue-driven degradation of collective coordination, documented in both physiological and team coordination research (Mohr et al., 2005; Duarte et al., 2013).

Crucially, fragmentation is asymmetric between teams in most match contexts. The team maintaining greater coherence in the collective motion field during the late game, characterised by lower V(t) and sustained positive G(t), tends to retain control of territory and tempo in the closing stages. This finding has implications for conditioning strategy, substitution timing, and late-game tactical design.

6. INTERPRETATION OF MOMENTUM

The concept of momentum in sport is widely invoked in commentary and coaching practice but has proven resistant to rigorous operationalisation. Existing approaches typically define momentum through scoring sequences (Vergin, 2000), possession streaks, or subjective ratings. Psychological momentum research has examined cognitive and emotional antecedents of perceived momentum (Bar-Eli, Avugos, & Raab, 2006; Gernigon, Briki, & Eykens, 2010), establishing that momentum is experienced as a dynamic, nonlinear phenomenon shaped by the history of performance within a contest.

The Action Gravity framework offers a complementary perspective grounded in observable collective dynamics rather than psychological inference. Momentum, within this framework, is operationally defined as a positive, sustained trajectory of AG(t) over a rolling window which is an emergent property arising from:

- Sustained directional coherence in the collective velocity field (high C_d(t)) over an extended passage.
- Progressive forward spatial dominance, evidenced by consistent positive G(t) relative to the field centre.
- Reduced opposition disruption, manifested as low and declining V(t) in the collective motion field.

This interpretation explains phenomena that are difficult to account for within event-based frameworks. The experience of pressure without points, in which a team exerts sustained attacking force without converting it to score, is explicable in terms of a high-AG trajectory that fails to achieve the positional threshold required for scoring opportunity. Sudden collapses in control, which appear discontinuous in event logs, are often preceded by gradual degradation in collective motion coherence that the AG signal makes visible.

The Action Gravity account of momentum also speaks to the psychological literature's concern with history dependence (Gernigon et al., 2010): the AG signal is inherently path-dependent, reflecting the accumulated directional history of the collective rather than treating each moment as independent. This provides a physical-behavioural correlate for the nonlinear, history-dependent dynamics of psychological momentum that Gernigon and colleagues identified.

7. PRACTICAL APPLICATIONS

7.1 Coaching and Performance Analysis

The Action Gravity framework provides practitioners with objective, theory-grounded metrics for evaluating team performance. Key applications include: identification of effective launch zones (strike vector analysis); detection of structural weaknesses (monitoring V(t) spikes during opposition phases); and objective evaluation of tactical execution (comparing intended structures against observed AG trajectories).

7.2 Automated Video Analysis

The MoP-based event detection capability (Step 8, Algorithm 1) enables automatic segmentation of match footage into classified passages by game phase, field position, and outcome. This supports construction of searchable, outcome-indexed clip libraries which has a significant practical advantage over manual tagging, which is labour-intensive and susceptible to annotation inconsistency (Carling et al., 2014).

7.3 Accessibility and Scalability

A distinctive feature of the framework is its reliance on video-derived data. Unlike GPS and optical tracking systems restricted to elite professional environments, video is available across the full spectrum of competition levels, from community sport to international fixtures. This enables application of advanced analytics at scales and in contexts currently unserved by existing approaches, including community rugby, school sport, and emerging competition contexts in regions without established tracking infrastructure.

8. EXTENSION TO OTHER SPORTS

The Action Gravity framework is inherently sport-agnostic. Its core requirement is a continuous spatial domain in which two collectives contest positional control through coordinated movement, a condition satisfied by sports like: association football, basketball, netball, field hockey and wheelchair rugby.

In association football, the framework complements existing pitch control and expected possession value modelling (Fernandez & Bornn, 2018) by adding an emphasis on collective motion coherence over time. The compression-expansion dynamic maps directly onto established tactical patterns such as high pressing and transitional counterattacking. In basketball, the high tempo and continuous positional transitions make the collective motion framework particularly apt; zone defence and fast-break offence represent canonical examples of compression and expansion respectively. In netball, the framework can be adapted to model spatial occupation within permitted zones and coherence of movement across the transverse axis.

Cross-sport application also raises the prospect of comparative analytics enabling systematic comparison of the structural dynamics of control across sports and potentially identifying universal principles of effective collective motion in competitive spatial systems.

9. FUTURE RESEARCH DIRECTIONS

9.1 Predictive Modelling

The collective motion field representation provides a natural foundation for predictive modelling of game outcomes. By framing match state as a continuous time-series of AG and CoM variables rather than a discrete event sequence, it becomes possible to model scoring probability, phase outcome, or territorial transition as a function of current and historical collective dynamics. This aligns with developments in multi-agent trajectory prediction (Alahi et al., 2016) and opens pathways to recurrent and attention-based neural architectures applied to sports outcome prediction.

9.2 Decision Quality Decomposition

A fundamental challenge in sports analytics is distinguishing good decisions with poor execution from poor decisions with favourable outcomes. The Action Gravity framework provides a pathway toward this decomposition by enabling comparison between the observed AG trajectory and the expected trajectory implied by a given tactical decision. A decision generating a theoretically coherent motion structure, even if it fails to produce a scoring outcome due to execution error, can be distinguished from one producing an incoherent motion structure that happens to yield an opportunistic score.

9.3 Optimisation and Quantum Formulations

The motion field representation admits reformulation as an optimisation problem: identifying collective motion configurations that maximise the AG signal subject to spatial constraints imposed by the opposition and field geometry. The combinatorial explosion of feasible multi-agent configurations under such constraints renders classical exhaustive search computationally intractable for real-time application.

This motivates exploration of Quadratic Unconstrained Binary Optimisation (QUBO) formulations, which provide a natural mapping from discrete action-selection problems to the objective function structure required for quantum annealing (Kadowaki & Nishimori, 1998). In a QUBO encoding, each binary variable represents a discrete movement decision for an individual player, and the quadratic objective captures pairwise interactions between player positions that contribute to or detract from the collective AG. Finding the optimal multi-player configuration then becomes equivalent to finding the minimum-energy state of an Ising Hamiltonian: a problem for which quantum annealing offers potential computational advantage as problem size scales.

9.4 Integration with Physiological Data

The video-derived framework can be extended through integration with player physiological data where available: including GPS accelerometry, heart rate, and metabolic load indicators. This integration would enable joint modelling of collective motion dynamics and individual physiological state, providing a more complete account of the mechanisms underlying late-game fragmentation and fatigue-related performance decline.

10. DISCUSSION

This paper has introduced the Action Gravity framework as a new approach to quantifying control in invasion sports. The framework makes three substantive contributions to the sports analytics literature.

First, it provides a conceptual reorientation from static spatial representation to dynamic motion-field analysis, capturing collective motion dynamics, coherence, direction, compression, and expansion, that are theoretically central to the generation of control but invisible to existing spatial methods.

Second, it operationalises this framework using video-derived positional data, making it accessible at all levels of competition. The algorithm presented in Section 4.3 provides a reproducible and extensible pipeline from raw video to interpretable motion-field signals.

Third, it provides a formal, model-based account of momentum as an emergent property of collective motion dynamics. This account generates testable predictions about the conditions under which momentum is generated, sustained, and disrupted. These predictions are grounded in the observable AG signal rather than post-hoc event attribution.

Dimension	Traditional Approach	Action Gravity
Spatial representation	Static location / heatmap	Dynamic motion field
Event model	Discrete event sequences	Continuous system state
Primary focus	Possession and territory	Structure and coherence
Data source	Proprietary tracking hardware	Match footage
Momentum	Scoring sequence / streaks	Emergent system property

Applicability	Elite environments only	Grassroots to elite

Table 1: Summary comparison of traditional approaches and the Action Gravity framework across key analytical dimensions.

The framework's explicit formal structure (Section 3.3) and reproducible algorithm (Algorithm 1) distinguish it from prior conceptual proposals and position it as both theoretically grounded and practically deployable. The connection to the psychological momentum literature (Bar-Eli et al., 2006; Gernigon et al., 2010) suggests that the AG signal may provide a behavioural correlate for phenomena previously studied only at the level of cognitive and emotional response.

11. LIMITATIONS

Player detection accuracy: The framework depends on the accuracy of the underlying object detection model. Occlusion, camera transitions, and variable image quality can introduce detection errors that propagate into the CoM and MoP signals. Robust smoothing (Step 4, Algorithm 1) mitigates but does not eliminate these effects.

Absence of explicit ball tracking: The baseline framework does not incorporate ball position data. Ball location is an important determinant of attacking structure and phase classification; its integration is a priority for future development.

Parameter estimation: The weighting parameters α, β, γ in the AG specification require calibration from annotated reference data. In the absence of such data, equal weighting provides a reasonable baseline, but optimal estimation may vary by sport, competition level, and tactical context.

Camera angle sensitivity: Homographic transformation quality depends on the availability of sufficient pitch landmarks within the video frame. Wide-angle and overhead video produce higher-quality transformations than tight sideline footage, placing some constraint on the range of applicable video sources.

These are recognised as current implementation constraints rather than fundamental theoretical barriers. Each has a clear development pathway, and the framework's modular architecture facilitates iterative improvement.

12. CONCLUSION

This paper has introduced Action Gravity, a dynamic collective motion framework for quantifying control in invasion sports. By reconceptualising play as a continuous motion field characterised by directional coherence, structural compression, and expansiodynamics. By formalising this as a computable scalar index $AG(t) = \alpha \cdot C_d(t) + \beta \cdot G(t) - \gamma \cdot V(t)$, the framework provides a theoretically grounded and empirically tractable approach to a long-standing analytical challenge.

The operationalisation through video-derived positional data is a deliberate design choice that extends the reach of advanced performance analytics to the full spectrum of competitive sport. Algorithm 1 provides a reproducible pipeline from raw video to interpretable motion-field outputs. The momentum reconceptualisation complements the psychological literature by providing a physical-behavioural correlate for the nonlinear, history-dependent dynamics that characterise perceived momentum in competition.

Action Gravity reframes invasion sport as a dynamic optimisation problem over collective motion, providing both a measurable construct of control and a foundation for predictive and computational advancement.

References

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 961–971.

Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of "hot hand" research: Review and critique. Psychology of Sport and Exercise, 7(6), 525–553. doi:10.1016/j.psychsport.2006.03.001

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014). Large-scale analysis of soccer matches using spatiotemporal tracking data. 2014 IEEE International Conference on Data Mining, 725–730. doi:10.1109/ICDM.2014.133

Bornn, L., Ward, P., & Norman, D. (2019). Training schedule confounds the relationship between acute:chronic workload ratio and injury. MIT Sloan Sports Analytics Conference.

Bracewell, P. J., Trowland, H. E., Wilson, J. K., & Tikhonov, A. (2020). Using machine vision to track the locus of play in rugby union. In R. Stefani & A. Schembri (Eds.), Proceedings of the 15th Australian Conference on Mathematics and Computers in Sports (pp. 57–60). ANZIAM MathSport.

Carling, C., Bloomfield, J., Nelsen, L., & Reilly, T. (2008). The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work-rate data. Sports Medicine, 38(10), 839–862.
Cintia, P., Rinzivillo, S., & Pappalardo, L. (2015). A network-based approach to evaluate the performance of football teams. Machine Learning and Data Mining for Sports Analytics Workshop, ECML/PKDD 2015.
Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S. (2015). Using network metrics in soccer: A macro-analysis. Journal of Human Kinetics, 45, 123–134. doi:10.1515/hukin-2015-0013
Duarte, R., Araújo, D., Correia, V., & Davids, K. (2012). Sports teams as superorganisms: Implications of sociobiological models of behaviour for research and practice in team sports performance analysis. Sports Medicine, 42(8), 633–642.
Duarte, R., Araújo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). Intra- and inter-group coordination patterns reveal collective behaviours of football players near the scoring zone. Human Movement Science, 31(6), 1639–1651. doi:10.1016/j.humov.2012.03.001
Fernandez, J., & Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. MIT Sloan Sports Analytics Conference.
Frencken, Wouter & Lemmink, Koen A.P.M. & Delleman, Nico & Visscher, Chris. (2011). Oscillations of centroid position and surface area of soccer teams in small-sided games. European Journal of Sport Science. 11. 215-223. 10.1080/17461391.2010.499967.
Gernigon, C., Briki, W., & Eykens, K. (2010). The dynamics of psychological momentum in sport: The role of ongoing history of performance patterns. Journal of Sport and Exercise Psychology, 32(3), 377–400.
Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. ACM Computing Surveys, 50(2), Article 22. doi:10.1145/3054132
Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. Journal of Sports Sciences, 20(10), 739–754.
Kadowaki, T., & Nishimori, H. (1998). Quantum annealing in the transverse Ising model. Physical Review E, 58(5), 5355–5363. doi:10.1103/PhysRevE.58.5355
Kelso, J. A. S. (1995). Dynamic patterns: The self-organization of brain and behavior. MIT Press.
Liebermann, D. G., Katz, L., Hughes, M. D., Bartlett, R. M., McClements, J., & Franks, I. M. (2002). Advances in the application of information technology to sport performance. Journal of Sports Sciences, 20(10), 755–769.
Mohr, M., Krustrup, P., & Bangsbo, J. (2005). Fatigue in soccer: A brief review. Journal of Sports Sciences, 23(6), 593–599. doi:10.1080/02640410400021286
Passos, P., Davids, K., Araújo, D., Paz, N., Minguens, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. Journal of Science and Medicine in Sport, 14(2), 170–176. doi:10.1016/j.jsams.2010.10.459
Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. SpringerPlus, 5, Article 1410. doi:10.1186/s40064-016-3108-2
Spearman, W. (2018). Beyond expected goals. MIT Sloan Sports Analytics Conference.
Vergin, R. C. (2000). Winning streaks in sports and the misperception of momentum. Journal of Sport Behavior, 23(2), 181–197.
Vicsek, T., & Zafeiris, A. (2012). Collective motion. Physics Reports, 517(3–4), 71–140. doi:10.1016/j.physrep.2012.03.004

DESIGNING A DATA AND MODELLING FRAMEWORK FOR RUCK OUTCOME ANALYSIS IN RUGBY UNION

Emmy Gauper[a,b], Paul J. Bracewell[a,d], F. S. Forbes[a,c], B. McGahan[a]
[a] *Dovetale Sport*
[b] *Victoria University of Wellington*
[c] *University of Southern Queensland*
[d] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Rucks are among the most frequent and strategically decisive events in rugby union, yet they remain comparatively under-modelled within performance analysis. This paper develops a structured data and modelling framework for analysing ruck outcomes, with particular emphasis on turnover prediction and tactical feasibility. We synthesise existing literature on ruck performance to identify key determinants of success, including player-entry numbers, arrival timing, spatial location, advantage-line progression, and ball-carrier technique. Building on this foundation, we propose a scalable ruck data schema designed for predictive modelling and optimisation. The framework restructures traditional long-format ruck data into a team-centric wide format, normalises field orientation relative to attacking direction, and formalises imbalance metrics and spatial encodings suitable for quantitative analysis. To demonstrate feasibility, we outline how logistic regression can be used to model ruck turnover probability and discuss extensions toward interaction terms and combinatorial structures. An interactive visualisation tool is introduced to support exploratory ruck analysis, linking player positions and join decisions to spatial context. Throughout, emphasis is placed on embedding tactical feasibility within modelling structures to ensure that statistically optimal recommendations remain operationally realistic.

Keywords: Event data modelling, Feature engineering, Predictive modelling, Quantum annealing

1. INTRODUCTION

Rucks are among the most common and consequential events in rugby union. Occurring after almost every tackle, the ruck determines which team retains possession of the ball and therefore governs the tempo, structure, and spatial progression of play. At the heart of every ruck is a real-time multi-agent decision problem: which players should commit to the contest, and when? These decisions, made under time pressure and with incomplete information, collectively determine not just individual ruck outcomes but the broader shape of a match. Despite their frequency and strategic importance, rucks remain one of the least analytically developed components of rugby union performance analysis, standing in sharp contrast to the modelling sophistication applied to areas such as set pieces or game-state analysis.

The relative scarcity of ruck modelling can be attributed to several interconnected challenges. First, rich ruck-level data has historically been difficult to obtain at scale. Unlike ball-in-hand statistics that are readily captured through broadcast feeds, ruck-level data typically requires manual video coding or bespoke computer vision pipelines. Second, the ruck is a multi-agent, spatially embedded event in which outcomes are shaped by the simultaneous decisions of many players, creating a combinatorial complexity that resists simple univariate analysis. Third, even when predictive models are constructed, the translation of statistically optimal recommendations into tactically actionable decisions is non-trivial.

This paper addresses these challenges by developing a structured framework for ruck outcome analysis. The framework encompasses three interconnected components: a principled data schema for ruck event capture, a feature engineering approach that operationalises domain knowledge into quantitative predictors, and a modelling architecture grounded in logistic regression with extensions toward combinatorial optimisation. The work is motivated by a growing body of sports analytics literature that demonstrates the practical value of event-data modelling in decision support for coaches and analysts.

The remainder of this paper is organised as follows. Section 2 reviews prior literature on ruck performance and the use of quantum annealing in sports analytics. Section 3 describes the data structure and feature engineering framework. Section 4 outlines the modelling approach. Section 5 discusses the interactive visualisation tool developed to support exploratory analysis. Section 6 presents results and discussion, and Section 7 concludes with directions for future work.

2. LITERATURE REVIEW

2.1 Ruck Performance in Rugby Union

A number of studies have examined ruck performance in rugby union, though the body of work remains comparatively sparse relative to other match events. O'Muilleoir (2024) used computer vision to automate the analysis of match footage, finding that rucks accounted for approximately 42% of all match events, confirming their status as the single most frequent event type in a match. Despite this frequency, most studies have not progressed beyond descriptive or chi-squared analysis of a small number of categorical variables.

Kraak, Coetzee, and Venter (2017) examined changes in match profiles following amendments to rugby union laws and found that in over 90% of rucks the attacking team maintained possession. This observation is critical for modelling: it confirms that turnover events are relatively rare, which has implications for class imbalance in predictive modelling. Chambers et al. (2019) used accelerometer and gyroscope data combined with 20-second video clips to classify ruck and tackle events using a random forest model, focusing on the period surrounding the moment of impact.

Hendricks et al. (2018; 2020) provided two of the most analytically detailed treatments of ruck performance. Their work identified that actively placing the ball and the direction in which the ball-carrier falls following a tackle significantly influence the probability of the attacking team retaining possession. They further found that having three to five defending players actively engaging in the ruck reduced the likelihood of the attacking team maintaining possession by 85% in the Championship context. Field position, match location, match status, and match quarter were each found to exert small but measurable effects.

Wheeler et al. (2013) used chi-squared tests to examine the effect of defensive ruck strategies, finding that early counter-ruck or jackal defensive approaches increased the probability of a turnover. The comprehensive thesis of Colomer et al. (2020) synthesised findings across numerous studies, noting that player ratios at the ruck, pitch location, and timing of ruck strategies all influenced ball possession outcomes. The systematic review by den Hollander et al. (2021) provides additional references for extending this literature.

2.2 Quantum Annealing in Sports Analytics

Quantum annealing is a meta-heuristic optimisation technique that leverages quantum mechanical phenomena, including quantum tunnelling, to search for the global minimum of an objective function. It is conceptually analogous to simulated annealing but operates on quantum hardware (QPUs) using qubits rather than classical bits. D-Wave Systems is the leading commercial provider of quantum annealing hardware, offering access via API.

To date, quantum annealing has not been applied to rugby union. Its applications in sports have been largely confined to scheduling problems (Pechler and Phillipson, 2025; Fujii and Matsui, 2021) and player selection optimisation (Iturrospe, 2021; Gunduz and Yilmaz, 2022; Toda et al., 2024). Pechler and Phillipson (2025) found that quantum annealing approaches did not yet outperform classical methods for scheduling problems, though this may change as hardware matures. The player selection approaches are of particular relevance to the ruck context, where the question of which players should join a forming ruck is a combinatorial optimisation problem.

The ruck join decision maps naturally onto a Binary Quadratic Model (BQM) or Quadratic Unconstrained Binary Optimisation (QUBO) formulation that is compatible with quantum annealing hardware. This connection motivates the modelling architecture proposed in this paper.

3. DATA AND FEATURE ENGINEERING FRAMEWORK

3.1 Raw Data Schema

The ruck data used in this study was collected by Dovetale Sport from video footage of seven sequential rucks during a match between the 1st XVs of St Patrick's College and Wellington College played in July, 2025. The data is structured in a long format, with each row corresponding to a specific player at a specific ruck. Six player positions were tracked per ruck: hooker (3), tight-head prop (4), both locks (5, 6), blind-side flanker (7), and number 8 (8). This yields 42 rows in total.

Table 1 summarises the key variables in the raw data schema. These include spatial coordinates for the ruck location and each player's position when the ruck formed, player join decisions, team-level ruck counts, and outcome variables for both the current and adjacent rucks.

Variable	Type	Description

Xruck, Yruck	Integer	Field coordinates (metres) where ruck was formed
XplayerWhenFormed, YplayerWhenFormed	Integer	Player's field position when the ruck formed
didJoin	Binary	Whether the player joined the ruck (Y/N)
ownTeamNum, oppTeamNum	Integer	Number of players from each team who entered the ruck
didWinRuck	Binary	Whether the player's team retained possession
AdvLinePrev, AdvLineCurr	Integer	X-coordinate of the previous and current ruck
frameStart, frameStable	Integer	Timing (seconds) of ruck formation and stabilisation
winNextEvent, XnextEvent, YnextEvent	Mixed	Future event indicators (excluded from predictive models)

Table 1: Summary of key ruck data variables

3.2 Data Restructuring Recommendations

The long-format structure of the raw data is appropriate for player-level exploratory analysis but suboptimal for ruck-level predictive modelling. We recommend restructuring to a wide format in which each row represents a single ruck, with player-specific information stored in position-coded columns (e.g., XPosition3WhenFormed rather than XplayerWhenFormed).

A further restructuring concerns team orientation. The raw data tracks a single team regardless of possession status. For modelling purposes, it is more appropriate to represent data from the perspective of the team in possession at the time of each ruck, since possession is the dominant determinant of outcome. Normalising field orientation relative to the attacking direction (so that the opponent's try line is always at coordinate 100) is also recommended to ensure consistency across matches and opposition.

3.3 Feature Engineering

Several feature engineering steps are proposed to operationalise domain knowledge into predictors suitable for logistic regression and quantum annealing models.

Spatial features: The raw X and Y coordinates of the ruck location can be combined into spatial zones (e.g., quadrants or 25-metre segments) for inclusion in quantum annealing models that require binary inputs. For logistic regression, continuous spatial coordinates are appropriate.

Advantage line progression: The advantage line concept, as operationalised by Hendricks et al. (2020), captures whether the ball-carrier gained or lost ground before being tackled. This can be approximated from AdvLinePrev and AdvLineCurr by computing the displacement and categorising it as 'Over', 'Equal', or 'Behind', using a body-length threshold of approximately 2 metres.

Player imbalance: The variables ownTeamNum and oppTeamNum can be combined into a numerical imbalance measure (e.g., ownTeamNum - oppTeamNum) or a binary flag indicating whether defenders outnumber attackers at the ruck. This is consistent with findings from Hendricks et al. (2018) who identified three to five defensive players as particularly influential.

Player distance and proximity: Combining XplayerWhenFormed and YplayerWhenFormed, a player's Euclidean distance from the ruck can be computed. This enables the construction of proximity-based imbalance metrics once opposing player coordinates are captured.

Temporal features: frameStart can be discretised into 10-minute match segments. FrameStable minus frameStart provides ruck stabilisation time, a proxy for the speed of the ruck contest, which has been shown to predict match outcomes in prior work.

Exclusion of future-event variables: Variables including winNextEvent, nextRuckStart, XnextEvent, and YnextEvent contain information about events that have not yet occurred at the time of the ruck. These are unsuitable as predictors in any real-time decision model and should be excluded.

4. MODELLING FRAMEWORK

4.1 Logistic Regression for Turnover Probability

Logistic regression is a natural starting point for ruck turnover modelling. It models the log-odds of a binary outcome (possession retained vs. turnover) as a linear function of predictors:

$$log[p(X) / (1 - p(X))] = \beta 0 + \beta 1X1 + \ldots + \beta pXp$$

where $p(X) = Pr(Y = 1 \mid X)$ is the probability of a turnover. Given the rarity of turnover events (Kraak et al., 2017, estimated retention rates above 90%), any logistic regression model applied to ruck data must account for

class imbalance. Approaches including oversampling the minority class, adjusting classification thresholds, or applying penalised regression may be warranted depending on the scale of the available dataset.

Recommended predictors for inclusion in an initial logistic regression model include: ruck spatial zone, player imbalance at the ruck, advantage line classification, match time segment, ruck stabilisation time, and team-specific effects. Variables such as ball-carrier falling direction, active ball placement, and ruck entry speed should be added to the data collection schema for future model iterations.

4.2 Extension to Combinatorial Optimisation

The ruck join decision is a combinatorial problem: given a set of available players at varying distances from a forming ruck, which subset should join? The number of feasible join combinations grows exponentially with squad size, reaching 2^15 = 32,768 possible configurations for a full 15-player squad, making brute-force enumeration impractical at match speed. This maps naturally onto a Binary Quadratic Model (BQM) in which each binary variable xi represents the join decision of player i, and the objective function encodes the predicted probability of winning the ruck as a function of these decisions.

In the QUBO formulation, the objective takes the form:

$$min\ \Sigma i\ ai\ xi + \Sigma i\ \Sigma j{>}i\ bij\ xi\ xj + c$$

where ai are linear coefficients derived from the logistic regression model, bij encode pairwise interaction effects between players, and c is an intercept. This formulation is directly compatible with D-Wave's quantum annealing hardware.

A hybrid solver approach, using D-Wave's Constrained Quadratic Model (CQM) framework, extends this to include continuous and integer-valued variables, enabling the incorporation of spatial coordinates and timing variables without requiring binarisation. A proof-of-concept implementation using open-source soccer data demonstrated that the logistic regression to QUBO pipeline is technically viable, yielding a ruck retention probability of approximately 93%. Tactical feasibility constraints must be incorporated to ensure that the optimiser does not recommend strategically impossible combinations.

5. INTERACTIVE VISUALISATION TOOL

To support exploratory analysis of ruck data without requiring video review, an interactive visualisation application was developed in R Shiny. The application allows the analyst to select a specific ruck by ID using a slider control and displays two linked views: a field plot showing the spatial position of the six tracked players at the time of ruck formation, and an overlay indicating the ruck location and whether each player joined. Players are colour-coded by their join decision (joined vs. did not join), enabling rapid visual assessment of which players were proximal to the ruck and chose not to join, and which were at greater distance but joined nonetheless. The ruck location is marked with a cross symbol. Sequential navigation through the slider allows the analyst to observe how the ruck contest evolves spatially across the match.

This tool fulfils several analytical functions. It provides a sanity check on data quality, enables pattern recognition at the exploratory stage, and serves as a communication tool for coaching staff who may be less familiar with tabular data representations. Future extensions could incorporate opposing team player positions, animated transitions between sequential rucks, and the overlay of model-derived join recommendations.

6. RESULTS AND DISCUSSION

6.1 Prototype Model Results

To demonstrate the end-to-end feasibility of the proposed pipeline, a logistic regression model was fitted to open-source soccer phase-of-play data from the 2024/2025 Australian A-League season (SkillCorner, 2025), using possession loss as the outcome variable. The soccer data was chosen as a structural analogue to the ruck context and serves purely as a methodological proof-of-concept; all domain-specific conclusions are drawn from rugby data.

Statistically significant predictors of possession loss included match minute, phase type (with 'chaotic' and 'set play' phases most strongly associated with turnover), and the x-coordinates of the phase start and end positions. The fitted model was subsequently converted to QUBO and CQM formats for hybrid and quantum annealing solver demonstration.

The hybrid solver identified the optimal strategy as a transition phase type commencing in the final minute near the opponent's goal, returning an objective value of -61.27 and a possession loss probability of approximately zero. While technically correct as an optimisation output, this recommendation is not

operationally feasible, highlighting the critical importance of embedding tactical feasibility constraints into the optimisation structure.

6.2 Implications for Ruck Modelling

The soccer proof-of-concept demonstrates that the logistic regression to QUBO pipeline is technically straightforward to construct and that hybrid solvers can handle the mixed-variable structures typical of ruck data. Several modelling challenges specific to the ruck context merit attention.

The class imbalance problem is acute: with turnover rates below 10% in most elite competitions, logistic regression models trained on imbalanced data will tend to assign near-zero turnover probabilities across the board. Addressing this requires deliberate sampling strategies or threshold calibration.

The combinatorial complexity of player join decisions grows rapidly with squad size. For a six-player candidate pool, there are 2^6 = 64 possible join combinations. In practice, with full squads of 15 players and real-time timing constraints, the combinatorial search space becomes large enough to motivate quantum or hybrid solvers.

Even from the small pilot dataset of seven rucks, preliminary patterns are evident. In rucks where the defending team outnumbered the attacking team by two or more players, the attacking team failed to retain possession in all observed instances. Conversely, in rucks where the attacking team achieved numerical parity or superiority, possession was retained in over 80% of instances. While these figures are based on a limited sample and should be interpreted with caution, they are directionally consistent with Hendricks et al. (2018).

The incorporation of interaction terms between player join decisions is a key modelling extension. If certain combinations of positions are particularly effective at winning rucks, these synergies should be encoded in the off-diagonal terms of the QUBO matrix. Estimating these interaction coefficients requires substantially more data than is currently available.

7. CONCLUSION

This paper has presented a structured data and modelling framework for ruck outcome analysis in rugby union. The framework addresses three linked challenges: data capture and schema design, feature engineering to operationalise domain knowledge, and predictive modelling with extensions toward combinatorial optimisation.

The key contributions are as follows. First, we propose a team-centric wide-format ruck data schema with normalised field orientation that is suitable for both logistic regression and quantum annealing models. Second, we specify a set of feature engineering transformations grounded in prior literature, including advantage-line classification, player imbalance metrics, proximity-based features, and temporal discretisation. Third, we outline a logistic regression modelling approach for turnover probability and demonstrate its extension to QUBO and CQM formats compatible with quantum annealing hardware. Fourth, we introduce an interactive R Shiny visualisation tool that links player positions and join decisions to spatial context.

The framework is designed to scale with richer tracking data. Future work will focus on three priorities: expanding the ruck dataset to cover multiple matches and teams; incorporating tactical feasibility constraints into the optimisation framework; and transitioning the proof-of-concept solver to D-Wave quantum hardware once API access is available. More broadly, this work reframes the ruck not as an isolated collision event, but as a structured decision system in which data, models, and optimisation can be combined to give coaches and players a genuine analytical advantage at the most frequent and consequential moment in a rugby match.

References

Chambers, R. M., Gabbett, T. J., Gupta, R., Josman, C., Bown, R., Stridgeon, P., & Cole, M. H. (2019). Automatic detection of one-on-one tackles and ruck events using microtechnology in rugby union. Journal of Science and Medicine in Sport, 22(7), 827–832.

Colomer, C. M., Pyne, D. B., Mooney, M., McKune, A., & Serpell, B. G. (2020). Performance analysis in rugby union: A critical systematic review. Sports Medicine–Open, 6(1), Article 4. https://doi.org/10.1186/s40798-019-0232-x

Den Hollander, S., Ponce, C., Lambert, M., Jones, B., & Hendricks, S. (2021). Tackle and ruck technical proficiency in rugby union and rugby league: A systematic scoping review. International Journal of Sports Science & Coaching, 16(2), 421–434. https://doi.org/10.1177/1747954120976943

D-Wave Quantum Inc. (2025). Leap service's hybrid solvers. https://docs.dwavequantum.com/en/latest/industrial_optimization/leap_hybrid.html

Fujii, K., & Matsui, T. (2023). Solving break minimization problems in mirrored double round-robin tournament with QUBO solver. arXiv:2307.00263.

Gündüz, S., & Yılmaz, İ. (2024). Harnessing the power of quantum computing to build an ideal team and line-up for the Euroleague fantasy challenge. International Journal of Innovative Approaches in Science Research, 8(2), 56–69. https://doi.org/10.29329/ijiasr.2024.1054.2

Hendricks, S., van Niekerk, T., Sin, D. W., Lambert, M., den Hollander, S., Brown, J., Maree, W., Treu, P., Till, K., & Jones, B. (2018). Technical determinants of tackle and ruck performance in international rugby union. Journal of Sports Sciences, 36(5), 522–528.

Hendricks, S., Till, K., den Hollander, S., Savage, T. N., Roberts, S. P., Tierney, G., Burger, N., Kerr, H., Kemp, S., Cross, M., et al. (2020). Consensus on a video analysis framework of descriptors and definitions by the Rugby Union Video Analysis Consensus Group. British Journal of Sports Medicine, 54(10), 566–572.

Iturrospe, A. (2021). Optimizing decision making for soccer line-up by a quantum annealer. arXiv:2112.13668.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

Kraak, W., Coetzee, F., & Venter, R. (2017). Analysis of the general match profile of international rugby union between 2007 and 2013. International Journal of Performance Analysis in Sport, 17(3), 303–318.

Ó Muilleoir, C. (2024). Advancing player modelling in rugby union: An evaluation of path analysis techniques [Dissertation/project report]. National College of Ireland.

Pechler, O., & Phillipson, F. (2025). Toward quantum annealing for multi-league sports scheduling. In S. Zielinski, G. Eichler, C. Erfurth, & G. Fahrnberger (Eds.), Innovations for Community Services: 25th International Conference, I4CS 2025, Munich, Germany, June 11–13, 2025, Proceedings (Communications in Computer and Information Science, Vol. 2513, pp. 159–188). Springer. https://doi.org/10.1007/978-3-031-94263-1_11

SkillCorner. (2025). Phases of play open data. GitHub. https://github.com/SkillCorner/opendata

Toda, K., Inoue, S., & Tobe, Y. (2024). Optimization of player-combinations in multiplayer games. In 2024 International Conference on Information Networking (ICOIN) (pp. 746–750). IEEE.

Wheeler, K., Lyons, K., & Harrinton, W. (2013). Effective defensive strategies at the ruck contest in rugby union. International Journal of Sports Science & Coaching, 8(3), 481–491. https://doi.org/10.1260/1747-9541.8.3.481

EVENT-LEVEL TURN DETECTION AND SPLIT-TIME ESTIMATION FROM SINGLE-CAMERA BRONCO TEST FOOTAGE USING ANCHOR-BASED STABILISATION AND PROBABILISTIC SMOOTHING

Swati S. [a,b], Bracewell P.J. [a,c], Joseph J. [a], McGahan B. [a]
[a] *Dovetale Sport*
[b] *Victoria University of Wellington*
[c] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Accurate split-time measurement in repeated shuttle tests such as the Bronco is central to performance monitoring in rugby and other team sports. While timing gates and wearable sensors provide high precision, their cost and deployment complexity limit routine use in field-based training environments. Single-camera video offers a scalable alternative; however, handheld recordings introduce camera motion, detection noise, and trajectory fragmentation that can degrade event-level timing accuracy.

This study presents an event-centric computational framework for automated turn detection and split-time estimation from single-camera Bronco Test footage. The pipeline integrates (i) anchor-based video stabilisation using a stationary reference object, (ii) Kalman filtering and Rauch–Tung–Striebel (RTS) smoothing for probabilistic trajectory refinement, (iii) velocity-based principal component analysis (PCA) to extract a one-dimensional motion signal aligned with the dominant running axis, and (iv) extremum detection with velocity sign-change refinement for precise turn localisation.

Three stabilisation strategies: template matching via normalised cross-correlation (NCC), deep learning–based anchor detection (YOLO), and a hybrid NCC–YOLO approach, are compared against manually annotated frame-level ground truth under both handheld and fixed-camera conditions. Results show that hybrid anchor tracking yields the lowest aggregate turn-detection error. Under handheld recording, mean absolute error in turning time is approximately 4 frames (≈0.13 s at 30 fps), compared to 3.4 frames (≈0.11 s) for fixed-camera footage. Far turns exhibit greater variability due to scale reduction and detection degradation at distance, while near turns remain consistently accurate.

These findings demonstrate that timing-accurate event detection is feasible from unconstrained handheld video when stabilisation and probabilistic smoothing are appropriately integrated. The framework provides a mathematically grounded, low-cost alternative for split-time analysis in applied sport settings and establishes a foundation for broader single-camera performance analytics in field-based testing.

Keywords: *Computer vision, trajectory estimation, video stabilisation, time-series smoothing*

1. INTRODUCTION

Video recordings play a central role in modern sport performance practice. Coaches and analysts routinely use footage to evaluate technique, monitor workload, and assess tactical decisions, while athletes review video to improve their own performance. With the widespread availability of smartphones and affordable cameras, video capture has become an integral tool not just in sports but across many industries. Similar video-based approaches are also applied in clinical rehabilitation and occupational health to study human movement. Across these domains, a common requirement is the automatic detection of important motion events from video. A motion event refers to a meaningful moment in a movement sequence, such as a sharp turn, change of direction, sudden stop, or sprint initiation. These events mark transitions between movement phases and are directly linked to performance demands. Accurate detection enables the measurement of split times, as well as technique evaluation, fatigue monitoring, and performance tracking. In high-performance sport, even small timing differences can be meaningful, making accurate and reliable automated detection an important practical objective. The automatic detection of motion events from video has wide practical relevance across many sports and movement-based activities. In cricket, identifying turning points during running between wickets provides

insight into sprint efficiency and decision-making, while in rugby and football, change-of-direction events are central to assessing agility, defensive positioning, and repeated high-intensity efforts (Thomas et al., 2017; Naik et al., 2022). Similar event-based analysis is relevant in athletics for sprint starts and lane turns, and in swimming for detecting wall turns and push-off phases (Mooney et al., 2016). Beyond sport, motion event detection is important in clinical rehabilitation (Muro-de-la-Herran et al., 2014) and occupational health (Faber et al., 2012).

Traditionally, performance measurement has relied on wearable sensors, timing gates, or multi-camera motion capture systems. Although these tools provide high accuracy, they are expensive, time-consuming to deploy, and not always available in real-world training environments. In addition, wearable sensors may interfere with natural movement or be impractical in competitive settings. Single-camera video offers a practical alternative, as almost every training session already has a camera available, making video-based analysis highly scalable and cost-effective.

However, extracting precise timing information from a single camera is challenging because real-world footage is often affected by handheld camera shake, motion blur during fast movements, partial and full occlusion, missed or unstable detections, and large changes in scale as athletes move farther from the camera. These factors introduce noise into measured athlete positions. Even small position errors can shift detected events by several frames, leading to noticeable errors in split-time estimates that may be unacceptable for performance analysis.

Principal Contributions: This paper makes three contributions: (1) an event-centric pipeline for turn detection from single-camera handheld video, evaluated explicitly for timing accuracy rather than identity tracking; (2) a hybrid anchor-stabilisation method that combines NCC template matching with YOLO-based detection, providing more reliable timing than either method alone; and (3) a velocity-informed extremum refinement approach that improves turn localisation beyond positional signal extrema alone.

Unlike most prior research, which relies on multi-camera systems or constrained recording environments, we evaluate a practical, video-only method that does not require specialised equipment or field calibration. The proposed approach does not require appearance-based re-identification and operates without homography estimation or field marking detection, making it applicable in unconstrained outdoor settings.

1.1 Related Work

The challenges of video-based performance analysis in unconstrained sport environments have been widely recognised, including occlusion, motion blur, scale variation, camera motion, and unstable viewpoints (Thomas et al., 2017; Naik et al., 2022). Together, these studies show that handheld footage with variable camera angles and uncontrolled lighting introduces substantial noise into tracking pipelines, directly affecting the reliability of downstream performance metrics.

Modern athlete tracking systems combine deep learning-based object detectors with multi-object tracking. Early work by Lu et al. (2013) demonstrated how geometry, motion cues, and tracklet association could preserve identity over time. Modern sports tracking pipelines commonly combine object detection, multi-object tracking, camera calibration or field registration, and sometimes appearance-based re-identification to preserve identity across frames and camera viewpoints (Redmon et al., 2016; Wojke et al., 2017; Luo et al., 2019; Zhang et al., 2022; Aharon et al., 2022; Pandya et al., 2023; Chen et al., 2025). These approaches are valuable for maintaining who is who over time, but their primary objective is identity preservation rather than precise localisation of when a motion event occurs. In practical sport footage, however, tracking pipelines can still produce fragmented trajectories because of missed detections, occlusion, camera motion, and scale variation. While existing methods demonstrate strong identity preservation, their primary objective is to maintain who is who over time rather than to precisely localise when motion events occur. Moreover, appearance-based re-identification is unreliable in practical training scenarios where athletes wear similar uniforms or appear small in the frame.

Camera motion further complicates single-camera analysis, necessitating video stabilisation. Classical stabilisation methods such as those proposed by Matsushita et al. (2006) and Grundmann et al. (2011) estimate camera motion by tracking generic background features. More recent learning-based approaches (Kerim et al., 2024) use neural networks to predict camera motion directly from image sequences. While these methods effectively reduce visual jitter, they are primarily designed to improve visual smoothness rather than to support precise event timing.

Studies using fiducial markers have demonstrated that tracking a stationary landmark can provide stable pose estimation over time (Risholm et al., 2021; Springer & Kyas, 2022). In the present study, we adopt a simpler anchor-based strategy: rather than estimating a full homography or camera pose, athlete coordinates are expressed relative to a persistent stationary object in the scene. In the present study, we adopt a simpler anchor-based strategy: rather than estimating a full homography or camera pose, athlete coordinates are expressed relative to a persistent stationary object in the scene.

Template matching and learned object detection provide complementary failure modes: NCC can be spatially precise when appearance is stable, while detector-based localisation can recover the anchor when the template match becomes unreliable. We therefore evaluate their combination empirically rather than presenting the hybrid strategy as an established prior method.

Various trajectory smoothing approaches have been applied to reduce noise in sports tracking. Probabilistic smoothing methods such as Kalman filtering and Rauch–Tung–Striebel (RTS) smoothing have been shown to reduce trajectory noise and improve motion stability (Fleck et al., 2023; Kalman, 1960; Lottefier et al., 2025; Rauch et al., 1965). Dimensionality reduction techniques, including PCA, are commonly used to project two-dimensional trajectories onto the dominant movement direction (Jolliffe, 2002). Signal smoothing techniques such as Savitzky–Golay filtering further reduce noise while preserving velocity peaks and turning points (Savitzky & Golay, 1964).

Overall, most prior research has focused on maintaining identity consistency and improving visual quality, rather than improving the accuracy of event-level timing estimates. Current pipelines are not explicitly designed to achieve the level of timing accuracy needed for reliable turn detection and split-time estimation from single-camera video. This gap motivates the need for an event-centric analysis framework that prioritises timing accuracy over long-term identity tracking.

1.2 The Bronco Test

The proposed methodology is evaluated using the Bronco Test, a repeated-sprint field test widely used in rugby and other team sports to assess high-intensity aerobic capacity and change-of-direction performance. The full Bronco consists of five consecutive shuttle sequences over 20 m, 40 m, and 60 m distances, requiring repeated acceleration, deceleration, and sharp turning. Because performance in the Bronco Test is highly sensitive to turning efficiency and split times, it provides a realistic and practically relevant benchmark for evaluating automated event detection. In this study, analysis focuses on one complete shuttle lap (20–40–60 m out-and-back), which captures the fundamental movement pattern repeated across all five laps. This setting enables systematic evaluation of automated turn detection and split-time estimation under both stable and handheld recording conditions, while clearly marked distance points provide ground-truth references for validating detected events.

2. DATA

This study analyses video data from two separate datasets: a primary handheld training recording used for methodological development, and a secondary set of paired stable and handheld recordings used to test the pipeline under controlled comparison conditions.

2.1 Primary Dataset

The primary dataset consists of a handheld video recording of multiple athletes performing the Bronco Test under typical training conditions, captured at 1080×1920 resolution (portrait orientation) at 30 frames per second (fps). The camera operator manually followed one designated athlete throughout the run, while other participants remained visible intermittently in the background. Analysis focused on the tracked athlete across one complete shuttle lap, capturing three far turning events at the 20 m, 40 m, and 60 m marks, and three near turning events at returns to the start line after 20 m, 40 m, and 60 m.

The footage contains natural camera motion, including position shifts, rotations, and scale changes as the tracked athlete moves toward and away from the camera. Partial and full occlusion occurred when athletes crossed paths, and the tracked athlete moved completely out of frame for approximately 50 frames (≈2.5% of all frames) during the sequence. A stationary pole served as a visual anchor for video stabilisation, remaining at least partially visible in most frames.

2.2 Secondary Dataset

The secondary dataset consists of simultaneous paired recordings of a single athlete performing the same shuttle sequence. One recording used a tripod-mounted stable camera that remained stationary throughout the trial, while the other used a handheld camera with deliberate camera motion. Both recordings were captured at 1920×1080 resolution (landscape orientation) at 30 fps, with a recording duration of approximately 56 s, capturing three far turning events and two near turning events.

In both secondary recordings, the athlete remained visible throughout the entire sequence with no occlusion events, and the anchor pole maintained continuous visibility across all frames. However, as the athlete moved farther from the camera toward the 60 m mark, their apparent size decreased substantially due to perspective, causing detection failures at maximum distance. The paired setup enables direct comparison of detection accuracy across camera conditions for identical movement sequences, isolating the impact of camera motion on timing precision.

2.3 Data Representation and Ground Truth

Video data were processed frame-by-frame, with athlete positions represented using bounding boxes defined by corner coordinates (x_1, y_1, x_2, y_2), from which centre coordinates are computed as $x^c = (x_1 + x_2)\times 2^{-1}$ and $y^c = (y_1 + y_2)\times 2^{-1}$. Manual annotation was used to establish ground-truth turn times and split times. Each video was reviewed frame-by-frame, and turning events were identified based on the moment when the athlete visibly changed running direction. A turn was defined as the frame corresponding to the instant of direction reversal, rather than the point of maximum displacement. All timestamps were stored in frame units and converted to seconds using the known frame rate (30 fps). Manual video annotation remains a common validation approach in applied sport analysis when event definitions are explicit and coding procedures are standardised (Bruce et al., 2022).

3. METHODOLOGY

We describe a pipeline for automated turn detection and split-time estimation in the Bronco Test under unconstrained conditions. The pipeline comprises eight sequential computational steps: (1) athlete detection, (2) anchor detection, (3) video stabilisation, (4) trajectory smoothing, (5) trajectory reconstruction and interpolation, (6) motion signal extraction via PCA, (7) signal smoothing, and (8) turn detection and split-time calculation. Each component is described in the following subsections.

3.1 Athlete Detection

Athlete detection provides frame-by-frame position estimation using the YOLOv9e model (Wang et al., 2024). YOLOv9e belongs to the YOLO family of single-stage object detectors, designed to provide a strong balance between localisation accuracy and computational efficiency. In this study, YOLOv9e was used with pretrained weights with no additional fine-tuning, as the detection task was limited to the 'person' class.

A minimum confidence threshold of 0.45 was applied to suppress low-probability bounding boxes. Evaluation over a representative 100-frame segment showed that this threshold most closely followed manually observed athlete counts, retaining valid detections while suppressing unstable or low-confidence bounding boxes. Because athletes wear identical uniforms and exhibit similar visual characteristics, appearance-based re-identification is assumed ineffective (Luo et al., 2019). Identity association relies exclusively on geometric continuity of bounding box positions across frames.

3.2 Anchor Detection

Anchor detection identifies a stationary reference object in the scene used to compensate for camera motion. In Bronco Test recordings, this anchor is typically a fixed field-side object such as a pole or post that remains static in the world but appears to move in the image due to handheld camera motion. The anchor is detected independently in each frame using three alternative strategies.

NCC template matching. Let A denote a reference anchor template extracted from an initial frame. For each candidate location (x, y) in frame t, the normalised cross-correlation (NCC) score $C^T(x, y)$ quantifies the similarity between the reference template and the candidate patch, with larger values indicating stronger visual correspondence. The optimal patch location is obtained as the location maximising $C^T(x, y)$. NCC is invariant to linear intensity changes and is therefore robust to moderate illumination variation (Lewis, 1995), but may degrade when the anchor undergoes significant scale changes, rotation, motion blur, or partial occlusion.

YOLO-based detection. As an alternative, the anchor is detected using a custom-trained YOLOv8s model that localises the stationary anchor directly from image appearance (Jocher, 2023). Unlike template matching, this approach does not rely on a fixed reference image and can accommodate moderate changes in anchor appearance. A low detection confidence threshold (0.05) was used to reduce missed detections when the anchor appeared small, blurred, or partially occluded.

Hybrid NCC–YOLO strategy. The hybrid strategy combines the strengths of both approaches. In the first frame, the anchor is initialised using YOLOv8s, and the corresponding image patch is stored as a reference template. For each new frame, NCC is applied first; if the similarity score exceeds 0.60, the NCC match is accepted as the anchor location. If the similarity falls below 0.60, the system switches to YOLO detection as a

backup (threshold 0.25). Detections that move more than 120 pixels between consecutive frames are rejected to prevent unintended anchor switching. The reference template is updated conservatively (similarity threshold 0.75) to adapt to gradual appearance changes without drifting toward an incorrect object.

3.3 Video Stabilisation

Handheld video recordings exhibit frame-to-frame camera motion that introduces unwanted variation in observed athlete positions. Because the objective of this study is accurate turn timing rather than visual stabilisation, the effect of camera motion is compensated directly in the extracted position signals using the stationary anchor. Camera-motion compensation is performed by expressing all athlete coordinates relative to the anchor location at each frame: $x^{stab}{}_1,t = x_1,t - x_a,t$, $y^{stab}{}_1,t = y_1,t - y_a,t$, with analogous transformations for the bounding box corner and centre coordinates. This transformation removes the effect of global translational camera motion while preserving the relative spatial geometry of athlete trajectories in image space. The procedure operates entirely in pixel coordinates and does not require camera calibration, physical field measurements, or homography estimation.

3.4 Trajectory Smoothing via Kalman Filtering and RTS Smoothing

Following video stabilisation, athlete detections remain noisy due to detection uncertainty, bounding box jitter, and residual localisation error. To obtain smooth and physically plausible motion signals, trajectory smoothing is applied independently to each track fragment using a state-space estimation framework.

The athlete state at frame t is represented by a 12-dimensional state vector containing centre coordinates, corner coordinates, and their corresponding velocity components. A linear state-space model with constant-velocity dynamics is assumed: $x^T = Fx^{T-1} + w^T$, where F is the state transition matrix and w^T is zero-mean Gaussian process noise with covariance Q. The constant-velocity model is a pragmatic short-time approximation of athlete motion between adjacent frames, with deviations (such as acceleration during turning) captured by the process noise rather than explicitly modelled.

The measurement model relates observed bounding box positions to the latent state: $z^T = Hx^T + v^T$, where H selects only the position components from the state vector and v^T represents measurement noise with diagonal covariance R. The Kalman filter proceeds recursively via prediction and update steps (Kalman, 1960). Because the entire video sequence is available offline, the Rauch–Tung–Striebel (RTS) backward smoother is subsequently applied to the filtered state estimates (Rauch et al., 1965). This adjusts each state estimate to be consistent with both past and future measurements, reducing temporal lag and suppressing local jitter more effectively than forward filtering alone.

3.5 Trajectory Reconstruction and Interpolation

Despite trajectory smoothing, athlete motion may be fragmented across multiple tracker IDs due to missed detections, short-term occlusions, or identity switches. Trajectory reconstruction focuses on merging these fragments and filling short temporal gaps to obtain a single continuous representation of the athlete's movement. Reconstruction is performed using geometric continuity of the stabilised and smoothed trajectories rather than appearance-based re-identification, using a look-ahead window to determine whether a lost track ID reappears shortly and evaluating candidate track IDs based on spatial proximity. Short missing segments are subsequently interpolated in the trajectory domain using neighbouring smoothed estimates. Longer gaps are left unfilled to avoid introducing implausible motion.

3.6 Motion Signal Extraction via PCA

Turn detection in the Bronco Test primarily depends on identifying reversals along the dominant direction of motion. The reconstructed trajectory is therefore reduced to a one-dimensional motion signal using PCA applied to RTS-smoothed velocity vectors, which provides a data-driven estimate of the dominant direction of travel (Jolliffe, 2002; Molenaar et al., 2013).

The velocity vectors are assembled into a matrix $V \in R^{(T\times2)}$, and PCA is applied to retain the first principal component $u \in R^2$ as the dominant motion axis. The RTS-smoothed position coordinates are then projected onto this axis to produce a scalar motion signal $s^T = [x^{c,T}\ y^{c,T}]\ u$, yielding a one-dimensional displacement signal indexed by frame t. In the analysed recordings, the estimated motion axis was approximately u = [0.75, −0.66] to u = [0.80, −0.60], indicating a diagonal direction in the image plane corresponding to the primary running path as viewed by the camera.

3.7 Signal Smoothing

The PCA-projected motion signal may still exhibit short-scale oscillations from residual frame-to-frame variability. A Savitzky–Golay filter is therefore applied, which performs local polynomial regression over a sliding window and is well suited for smoothing motion signals without distorting key geometric features such as slope and curvature (Savitzky & Golay, 1964).

A two-pass smoothing strategy balances overall trajectory stability with local precision around turning points. The first pass applies a wider window (121 frames, ≈4 s at 30 fps) to reduce long-range oscillations across each shuttle segment, ensuring that displacement remains smooth and monotonic between turns. The second pass applies a shorter window (61 frames, ≈2 s) to remove smaller residual fluctuations near turning points while preserving local curvature required for accurate extremum detection.

3.8 Turn Detection and Split-Time Calculation

Turning events correspond to reversals in the athlete's direction of motion along the one-dimensional PCA-projected signal. Candidate turning points are first identified as local extrema of the smoothed motion signal using a prominence-based criterion to suppress spurious peaks. Each detected extremum is refined using quadratic interpolation over the surrounding three frames to estimate peak location more precisely.

Detected extrema are classified as far or near turning points based on their relative displacement magnitude. To enforce physically meaningful alternation between outward and return running phases, consecutive extrema assigned the same label are collapsed, retaining only the first occurrence.

Although extrema provide a coarse estimate of turning locations, the true turning instant corresponds more closely to the moment at which the athlete's velocity along the running direction changes sign. Candidate extrema are therefore adjusted using RTS-smoothed velocity signals within a local temporal window. When multiple candidate sign changes are available, the refinement is selected based on the dominant PCA motion axis. If no valid sign change is detected within a bounded neighbourhood, the original positional extremum is retained.

Once refined turning points are obtained, split times are computed by converting frame indices to time using the video frame rate and calculating the duration of outward and return segments as $Out_i = (t^{far}_i - t^{near}_i) / FPS$ and $Back_i = (t^{near}_{i+1} - t^{far}_i) / FPS$.

4. RESULTS

Results are presented in four parts. The first two subsections use the primary dataset to assess the feasibility of recovering interpretable motion from handheld video and to compare anchor stabilisation strategies. The final two subsections use the secondary paired dataset to validate the motion representation and to quantify turn and split-time accuracy relative to manual annotations.

4.1 Stabilisation of Handheld Video

We first assess whether anchor-based stabilisation is able to suppress camera-induced motion in handheld video sufficiently to reveal the underlying forward–backward movement of the athlete. The raw trajectory exhibits large-amplitude fluctuations and long-term drift dominated by handheld camera movement, obscuring the underlying forward–backward structure and making direct interpretation of motion phases difficult.

After anchor-based stabilisation, the dominant camera-induced drift is removed, and the stabilised trajectory shows a repeated rise-and-fall pattern corresponding to outward and return running phases. Applying Kalman filtering followed by RTS smoothing further suppresses residual noise while preserving the temporal structure of the motion. The PCA-projected signal exhibits a clear oscillatory structure, with large-amplitude excursions corresponding to outward running phases and intervening troughs corresponding to return phases. After Savitzky–Golay smoothing, the consistent ordering and spacing of smoothed peaks and troughs matches the expected out-and-back shuttle structure of the Bronco Test, demonstrating that the dominant direction of motion has been successfully isolated despite substantial camera motion in the original footage.

4.2 Comparison of Stabilisation Methods

Frame-wise anchor trajectories were compared across three stabilisation strategies. Across most of the sequence, the NCC-based tracker produces a smooth and slowly varying anchor trajectory, indicating stable template

matching and consistent anchor localisation. This behaviour is reflected in the early far-turn results in Table 1, where NCC detection errors remain within ±7 frames for the first two far turns.

At the final far turn, however, NCC exhibits a substantially larger timing error (−37 frames). This error may arise from a small but consistent bias in the estimated anchor position caused by subtle changes in the image that are not perfectly captured by simple translation-based template shifts. Even small systematic offsets can be amplified during camera-motion compensation and motion-signal extraction, leading to a shifted estimate of the turn location.

In contrast, the YOLO-based detector exhibits several abrupt discontinuities in both anchor coordinates, corresponding to incorrect or missed anchor detections at individual frames. Unlike the NCC bias, these errors occur at isolated frames rather than persisting over time. As a result, YOLO achieves a much smaller error at the same far turn (±2 frames) despite exhibiting visually noisier anchor trajectories.

The hybrid method combines the strengths of both approaches. It follows the smooth temporal profile of NCC when template matching is reliable, while selectively reverting to YOLO detections to correct systematic bias when NCC assumptions are violated. The hybrid strategy avoids the large impulsive outliers observed in the YOLO-based tracker while maintaining a smooth anchor trajectory comparable to NCC.

Table 1. Comparison of detected turning frames across anchor stabilisation methods. Errors are reported in frames relative to manually annotated ground truth (GT).

Turn	Type	GT Frame	NCC Error	YOLO Error	Hybrid Error
1	FAR	177	+4	+7	+6
2	NEAR	330	−3	−3	−4
3	FAR	620	−7	−4	−4
4	NEAR	901	−26	−26	−26
5	FAR	1370	−37	+2	+1
6	NEAR	1842	+1	+1	0

A notable exception occurs at the second near turn (Turn 4), where all three stabilisation methods exhibit an identical −26 frame error. This error is not attributable to anchor tracking or stabilisation quality but coincides with a prolonged interval in which the athlete exits the camera field of view for approximately 50 consecutive frames. Because the true turning event occurs during or immediately adjacent to this missing-data interval, the precise timing cannot be recovered from video alone.

Table 2. Mean Absolute Error (MAE) in detected turning frames and equivalent timing error (seconds, at 30 fps)

Method	FAR MAE (frames)	FAR MAE (s)	NEAR MAE (frames)	NEAR MAE (s)	Overall MAE (frames)	Overall MAE (s)
NCC	16.00	0.533	10.33	0.344	13.00	0.433
YOLO	4.33	0.144	10.33	0.344	7.17	0.239
Hybrid	3.67	0.122	10.00	0.333	6.83	0.228

Table 2 shows that the hybrid method yields the lowest aggregate timing error across all events, with YOLO performing comparably and NCC exhibiting the largest errors due to its far-turn failure. Near-turn errors are similar across methods, reflecting the dominant influence of athlete visibility and occlusion rather than anchor stabilisation quality. Together, these results identify the hybrid approach as the most reliable stabilisation strategy for event-level timing in single-camera Bronco Test footage.

4.3 Accuracy of Turn Detection Against Manual Annotations

Detected turn frames were quantitatively compared against manually annotated ground truth for one participant completing a full shuttle lap (20–40–60 m) under stable and handheld camera conditions. Tables 3 and 4 quantify the temporal accuracy of automatically detected turning events relative to manual annotations.

Table 3. Frame-wise comparison of detected turns against manual ground truth for stable and handheld conditions.

Turn	Type	Stable GT	Stable Det.	Stable Err.	Handheld GT	Handheld Det.	Handheld Err.
1	FAR	169	172	+3	170	173	+3
2	FAR	610	607	−3	612	616	+4
3	FAR	1301	1308	+7	1303	1311	+8

Table 4. Absolute Mean Error (MAE) for stable and handheld camera conditions.

Condition	FAR MAE (frames)	FAR MAE (s)	NEAR MAE (frames)	NEAR MAE (s)	Overall MAE (frames)	Overall MAE (s)
Stable	4.33	0.144	2.0	0.067	3.4	0.113
Handheld	5.0	0.167	2.5	0.083	4.0	0.133

At the individual event level (Table 3), detected turn frames are generally close to ground truth across both camera setups, with errors remaining within ±8 frames for all events. Far turns exhibit larger variability than near turns under both conditions. In particular, the third far turn (60 m) shows the largest deviations, reflecting reduced detection reliability at long distances where the athlete appears small and blurred. Near turns, occurring closer to the camera, are detected with smaller and more consistent errors.

Aggregated results (Table 4) show that the stable camera condition achieves lower error overall. The overall MAE increases from 3.4 frames (0.113 s) under stable conditions to 4.0 frames (0.133 s) under handheld conditions. These results indicate a modest but consistent reduction in temporal accuracy when using handheld video, particularly for far turning events. Nevertheless, the magnitude of the error remains relatively small in absolute terms, demonstrating that the proposed pipeline is capable of recovering turn timings from handheld footage with accuracy comparable to that obtained under stable recording conditions.

5. DISCUSSION

This study investigated whether single-camera, handheld video can support reliable event-level timing through anchor-based stabilisation and motion-axis projection, using the Bronco Test as a representative application. The results demonstrate that interpretable forward–backward motion can be recovered from handheld footage and that turning events can be detected with mean absolute errors on the order of 0.1–0.2 s when stabilisation, smoothing, and signal refinement are combined appropriately.

5.1 Key Findings

Three principal insights emerge from this work. First, reference-based stabilisation is essential for suppressing camera-induced motion in handheld video. Without adjustment, raw athlete trajectories are dominated by camera movement, partially masking the true underlying movement structure. Anchor-based stabilisation, followed by probabilistic smoothing, can effectively recover a stable motion signal that reflects athlete dynamics rather than camera artefacts.

Second, dimensionality reduction via PCA applied to smoothed athlete trajectories reliably preserves the ordering and phase structure of shuttle running, enabling consistent identification of outward and return segments even under handheld conditions. However, PCA-based position signals encode where turns occur in the movement cycle rather than exactly when they occur, and velocity-based refinement is necessary for precise timing.

Third, stabilisation quality has a measurable impact on event-level timing accuracy. Hybrid anchor tracking, which combines the temporal smoothness of NCC template matching with the robustness of YOLO-based detection, yields the most reliable performance across varying distances. Nevertheless, even with optimal stabilisation, distant turning events remain more challenging to localise than near turns due to scale reduction, perspective effects, and reduced detection reliability.

5.2 Limitations

Several limitations of the current approach should be acknowledged. First, the quantitative evaluation is based on a limited number of video sequences. The paired dataset used for accuracy assessment consists of only two

recordings and a single completed shuttle lap. While this is sufficient to demonstrate feasibility and to isolate the effect of camera motion, it does not allow conclusions regarding variability across repeated shuttles, fatigue-related changes in running mechanics, or between-session consistency. Larger datasets would be required to assess robustness; confidence intervals and statistical significance testing across conditions are identified as a priority for future work with expanded datasets.

Second, the accuracy of turn localisation decreases when the athlete is far from the camera and detection quality deteriorates. In the final far turns, extreme motion blur and reduced athlete visibility led to missed detections and greater reliance on trajectory interpolation.

Third, split-time estimation is based on velocity-derived turn detection rather than explicit detection of an athlete crossing predefined spatial markers. While this enables event inference from unconstrained video without requiring field instrumentation, it means that the first and final legs of a shuttle cannot be independently estimated within this formulation.

Fourth, the pipeline relies on anchor-based stabilisation and therefore assumes the availability of a visible, stationary reference object for most of the recording. In many outdoor sports environments, suitable anchors are commonly present, but indoor venues or environments lacking fixed reference structures may require adaptation.

Finally, the current implementation assumes fixed camera intrinsics and constant video resolution. The method does not explicitly handle zoom operations, dynamic focal length changes, or substantial camera repositioning. Extending the framework to incorporate scale-aware stabilisation or explicit camera parameter estimation would be necessary for fully unconstrained handheld recording scenarios.

5.3 Directions for Future Work

Several avenues for future work follow naturally from this research. The proposed pipeline should be evaluated on larger and more diverse datasets, including multiple athletes and repeated shuttle laps, to allow assessment of inter-athlete variability, fatigue effects, and consistency of turn timing across repeated trials. Such expanded data would also enable confidence intervals, variance analysis, and paired significance testing.

Improving robustness when the athlete is far from the camera remains an important area for future work. Future improvements could make the interpolation process more adaptive to detection quality, assigning lower trust to frames with low confidence scores or adjusting smoothing strength accordingly. Explicitly modelling camera zoom and focal-length variation would enable more robust operation under unconstrained handheld recordings.

Extending the framework to other repeated-sprint tests and change-of-direction tasks would demonstrate broader applicability beyond the Bronco Test and establish the method as a general-purpose tool for single-camera performance analytics in field-based testing.

5.4 Practical Implications

The results indicate that single-camera handheld video can support meaningful event-level analysis with relatively modest error, enabling timing-grade analysis without wearables or timing gates. For many applied contexts, including routine training monitoring and split-time estimation, the observed accuracy is likely sufficient to inform performance assessment without specialised equipment. For applications requiring very high temporal precision, such as fine-grained biomechanical analysis or elite performance diagnostics, the residual errors at distant turning points may remain limiting.

Overall, this work demonstrates that anchor-based stabilisation and motion-axis projection provide a viable foundation for extracting interpretable motion and timing information from handheld video, while also clarifying the conditions under which such approaches succeed and where their limits lie.

References

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2206.14651.

Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2011). Estimation with Applications to Tracking and Navigation (2nd ed.). Wiley.

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Bruce, L., Dwyer, D. B., Fox, A., & Cè, E. (2022). Reliability of live and video-based coding in netball using the NetballStats application. PLOS ONE, 17(6), e0269330.

Chen, H.-Q., Liao, C.-C., Sun, Y.-H., Lin, C.-K., & Tseng, Y.-C. (2025). FieldMOT: A Field-Registered Multi-Object Tracking for Sports Videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Faber, G. S., Kingma, I., & van Dieën, J. H. (2012). Using biomechanical exposure assessment tools to evaluate work-related musculoskeletal disorder risk: A systematic review. Applied Ergonomics, 43(2), 354–363.

Fleck, T., Zipfl, M., & Zöllner, J. M. (2023). Semi-Automatic Ground Truth Trajectory Estimation and Smoothing Using Roadside Cameras. In IEEE International Conference on Intelligent Transportation Systems (pp. 4577–4583).
Grundmann, M., Kwatra, V., & Essa, I. (2011). Auto-directed video stabilization with robust L1 optimal camera paths. ACM Transactions on Graphics, 30(4).
Harris, C. R., et al. (2020). Array programming with NumPy. Nature, 585, 357–362.
Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95.
Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8.
Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME—Journal of Basic Engineering, 82(1), 35–45.
Kerim, A., Ramos, W. L. S., Marcolino, L. S., Nascimento, E. R., & Jiang, R. (2024). Leveraging Synthetic Data to Learn Video Stabilization Under Adverse Conditions. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.
Lewis, J. P. (1995). Fast normalized cross-correlation. Vision Interface, 120–123.
Lottefier, J., Van den Broeck, P., & Van Nimmen, K. (2025). Vision-Based Trajectory Reconstruction in Human Activities: Methodology and Application. Sensors, 25(24), 7577.
Lu, W.-L., Ting, J.-A., Little, J. J., & Murphy, K. P. (2013). Learning to track and identify players from broadcast sports videos. IEEE Transactions on Pattern Analysis and Machine Intelligence.
Luo, H., Gu, Y., Liao, X., Lai, S., & Jiang, W. (2019). A strong baseline and batch normalization neck for deep person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
Matsushita, Y., Ofek, E., Ge, W., Tang, X., & Shum, H.-Y. (2006). Full-frame video stabilization with motion inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
McKinney, W. (2011). pandas: A foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing.
Molenaar, P. C. M., Wang, Z., & Newell, K. M. (2013). Compressing movement information via principal component analysis: Contrasting outcomes from the time and frequency domains. Human Movement Science, 32(6), 1495–1511.
Mooney, R., Corley, G., Godfrey, A., Osborough, C., Quinlan, L. R., & ÓLaighin, G. (2016). The use of wearable sensors and video analysis in athletics. Sports Medicine, 46(5), 663–681.
Muro-de-la-Herran, A., Garcia-Zapirain, B., & Mendez-Zorrilla, A. (2014). Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. Sensors, 14(2), 3362–3394.
Naik, B. T., Hashmi, M. F., & Bokde, N. D. (2022). A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. Applied Sciences, 12(9), 4429.
Pandya, Y., Nandy, K., & Agarwal, S. (2023). Homography Based Player Identification in Live Sports. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
Rauch, H. E., Tung, F., & Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. AIAA Journal, 3(8), 1445–1450.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779–788).
Risholm, P., et al. (2021). Underwater marker-based pose-estimation with associated uncertainty. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.
Särkkä, S. (2013). Bayesian Filtering and Smoothing. Cambridge University Press.
Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 36(8), 1627–1639.
Springer, J., & Kyas, M. (2022). Autonomous Drone Landing with Fiducial Markers and a Gimbal-Mounted Camera for Active Tracking. In Proceedings of the 6th IEEE International Conference on Robotic Computing (pp. 243–247).
Szeliski, R. (2011). Computer Vision: Algorithms and Applications. Springer.
Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports: Current applications and research topics. Computer Vision and Image Understanding, 159, 3–18.
Virtanen, P., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261–272.
Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. In European Conference on Computer Vision.
Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In IEEE International Conference on Image Processing (pp. 3645–3649).
Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In European Conference on Computer Vision.

POSSESSION WITHOUT THE BALL: LATENT-STATE INFERENCE FROM COLLECTIVE MOVEMENT IN RUGBY UNION

Wei, Q., [a,b] Bracewell, P.J., [a,c] McGahan, B., [a] Joseph, J. [a]
[a] *Dovetale Sport*
[b] *Victoria University of Wellington*
[c] *Corresponding author: paul.bracewell@dovetalesport.com*

Abstract

Ball possession is a fundamental descriptor of tactical control in Rugby Union, yet it is typically obtained through manual annotation or specialised ball-tracking systems that are expensive and often unavailable outside elite environments. This study establishes that possession is identifiable as a latent game state using player tracking data alone, providing a proof-of-concept framework for automated possession inference without ball tracking.

Using frame-level two-dimensional coordinates extracted from broadcast footage, interpretable spatio-temporal descriptors of collective team behaviour are constructed, including Centre of Play (CoP), spatial dispersion, Movement of Play (MoP), movement dynamics, and temporally smoothed trajectory features. Possession labels are manually annotated for supervision; no ball-tracking information is used in model construction.

Random Forest classifiers are trained under both global and region-specific modelling frameworks. The global model achieves an accuracy of 76.2% and an AUC of 0.836, with repeated random-split validation yielding a mean AUC of 0.867 (±0.057), demonstrating stable discriminative performance. Region-specific models reveal substantial spatial heterogeneity, with the central interaction zone achieving an AUC of 0.964, indicating that possession signals are strongest in territorially structured areas.

These findings reframe possession as an emergent property of coordinated team movement, a latent state that is identifiable without observing the ball, and provide a computationally scalable and deployable framework for automated performance analysis in rugby and other invasion sports where ball trajectories are unavailable.

Keywords: Machine learning classification; Spatial behaviour modelling; Spatio-temporal feature extraction

1. INTRODUCTION

Rugby Union is a highly dynamic team-based invasion sport. Unlike football or basketball, where the ball remains visible for most of play, Rugby Union contains extended periods of ball occlusion: particularly during scrums, rucks, mauls, and tackles. Although modern ball-tracking systems can capture ball trajectories with high precision, their implementation remains technically complex and generally restricted to elite professional settings. As a result, most practical datasets provide player tracking data only, leaving ball position and possession unobserved.

Player tracking data encode underlying tactical structure: attacking teams tend to advance, create width, and support depth, whereas defensive teams compress space and reorganise their formation. Prior research in invasion sports has shown that collective positional structure and movement dynamics reflect tactical state and game context even without ball information (Gudmundsson & Horton, 2017; Low et al., 2020). If these patterns can be quantified, possession may be inferred without explicit ball tracking.

In current rugby performance analysis practice, performance is primarily quantified using manually coded event-based indicators derived from annotation software (Colomer et al., 2020). Possession, in particular, is typically annotated manually; a process that is labour-intensive, difficult to scale, and subject to analyst variability. An automated method for inferring possession from player tracking data alone would substantially improve analytical efficiency and scalability.

Despite extensive research in football, basketball, and similar sports, to the authors' knowledge no published study has demonstrated automated possession classification in Rugby Union using player tracking data alone. The primary aim of this study is therefore to develop a proof-of-concept framework for inferring ball possession using frame-level two-dimensional player coordinates, treating possession as a latent game state consistent with established latent-state modelling frameworks (Rabiner, 1989). This study makes three specific contributions:

- First empirical demonstration that possession is identifiable as a latent state from player tracking data alone, without any ball-tracking information;

- Reframing of possession as a latent spatio-temporal state; an emergent property of coordinated collective movement rather than a ball-centric event;
- Empirical evidence of spatial heterogeneity in possession signals across pitch regions, with region-specific models substantially outperforming a global baseline in tactically structured areas.

Random Forest classifiers were selected as a methodological baseline due to their interpretability, robustness to high-dimensional feature spaces, and suitability for establishing the existence of a signal prior to exploring sequence-based alternatives such as Hidden Markov Models or Long Short-Term Memory networks. The remainder of the paper is organised as follows. Section 2 reviews relevant literature. Section 3 describes the data. Section 4 details the methodology. Section 5 presents results. Section 6 discusses findings and limitations, and Section 7 concludes.

2. LITERATURE REVIEW

2.1 Spatio-temporal Analysis Using Player Tracking Data

The increasing availability of player tracking data has driven substantial progress in the analysis of team invasion sports (Gudmundsson & Horton, 2017). Modern tracking systems provide continuous observations of player movements throughout matches, enabling the analysis of evolving tactical behaviour rather than relying on discrete events or summary statistics. Tracking data thus provide a richer and more objective representation of match dynamics (Kovalchik, 2023).

Bartlett et al. (2012) analysed team coordination patterns in elite soccer using player movement trajectories, introducing team centroids and measures of spatial dispersion to characterise collective behaviour. Their findings showed that team dispersion varies systematically across phases of play, with more compact spatial organisation observed during defensive phases and more expansive structures during successful attacking phases providing early empirical evidence that possession-related information may be implicitly encoded in team-level spatial organisation. Systematic reviews have since identified spatial dispersion, centroid dynamics, and coordination synchrony as core feature families for characterising team tactics using tracking data (Low et al., 2020).

2.2 Event Detection and Latent Game State Inference

In football, automated pipelines have demonstrated that match events and possession can be detected from tracking data without manual annotation, indicating that possession-related information is implicitly encoded in coordinated spatio-temporal movement patterns (Vidal-Codina et al., 2022). Similarly, prior studies in rugby have shown that key match events such as tackles, rucks, and scrums can be identified using player tracking or wearable sensor data (Kelly et al., 2012; Chambers et al., 2019). Since possession transitions in Rugby Union frequently occur in conjunction with such events, this body of work motivates the use of event information as contextual cues in possession modelling.

2.3 Challenges of Possession Inference in Rugby Union

In contrast to much of the football literature, where possession is often defined at the individual, ball-centric level using player–ball proximity (Link & Hoernig, 2017), Rugby Union typically lacks reliable ball tracking. Rugby Union is further characterised by frequent physical collisions, structured restarts, rapid turnovers, and extended periods of ball occlusion during contact phases. Player formations frequently become highly congested, and possession may change without clear spatial separation between teams. A recent systematic review confirms that most existing work concentrates on possession outcomes rather than automated inference of possession states (Guerrero-Esteban et al., 2023), leaving a clear methodological gap that the present study addresses.

3. DATA

3.1 Source and Tracking Extraction

Player tracking data were extracted from broadcast video footage of a single competitive Rugby Union match using a computer vision–based optical tracking approach applied to footage captured by a pan–tilt–zoom (PTZ) camera. Analysis was restricted to the first half of the match to ensure consistency in attacking direction and spatial interpretation of positional features. The raw dataset comprised 46,422 observations of two-dimensional (x, y) player coordinates across 71,000 video frames. Following removal of non-player entities, the final processed dataset contained 46,136 rows of player tracking data.

For modelling, player positions from both teams were aggregated into a single collective formation without conditioning on team identity, ensuring that analysis focused on general possession movement patterns rather

than team-specific styles. Possession labels were manually annotated from the original video footage at 20-second intervals and temporally aligned with tracking frames to serve as the ground-truth binary response variable.

3.2 Coordinate Transformation and Summary Statistics

Raw video-based coordinates were normalised using Min-Max Normalisation, then linearly rescaled to a pitch-aligned coordinate system (100 m × 70 m) to facilitate interpretable spatial comparisons. Table 1 summarises the descriptive statistics of the core positional variables prior to feature smoothing.

Feature	Mean	S.D.	Min	Max	Skewness	Kurtosis
mean x (m)	48.84	26.17	5.85	95.64	0.05	−1.21
mean y (m)	40.44	10.64	10.16	65.12	−0.64	−0.20
sd x (m)	7.49	2.67	0.00	22.02	0.42	0.39
sd y (m)	12.28	4.22	0.71	24.94	−0.61	−0.04

Table 1: Descriptive statistics of positional spatial features.

4. METHODOLOGY

4.1 Framework Overview

The conceptual pipeline is as follows: Player Tracking → Feature Extraction (CoP, MoP, Dispersion) → Smoothing →Random Forest →Possession Probability. Player tracking coordinates are aggregated into frame-level spatial descriptors, smoothed to suppress measurement noise, and used as inputs to a supervised Random Forest classifier that outputs a probabilistic possession estimate for each frame.

4.2 Feature Derivation

Positional structure. For each frame, team-level spatial descriptors were computed by aggregating the positions of all observed players, treating both teams symmetrically. The mean horizontal and vertical positions, the Centre of Play (CoP), represent the spatial centroid of collective player movement (Bartlett et al., 2012). Spatial dispersion was quantified using the standard deviations of player positions in the horizontal and vertical directions (sd_x, sd_y), capturing the compactness or spread of the collective formation. Higher-order distributional moments (skewness and kurtosis) were computed to capture asymmetry and tail behaviour in the spatial distribution of players, reflecting formation imbalance and spatial concentration during breakdown phases.

Movement of Play (MoP). A metric termed the Movement of Play (MoP) was derived as the square root of the sum of spatial variances of player positions in the horizontal and vertical directions at each frame. This metric reflects the overall spatial spread and irregularity of the collective formation, with larger values indicating less stable or more dispersed movement patterns.

Movement dynamics. Collective movement dynamics were characterised using first-order differences in the CoP. Based on centroid displacement, collective movement speed (magnitude of CoP displacement per unit time) and direction of movement (angle of displacement relative to the horizontal axis) were derived.

Trajectory smoothing. To reduce measurement noise inherent in video-based tracking, smoothing was applied to the time series of CoP positions and the MoP metric. An adaptive smoothing span, calibrated to approximately 18 seconds of match footage, was employed. This effectively filters high-frequency noise while preserving lower-frequency tactical trends. All subsequent modelling was performed using these smoothed features. Figure 1 illustrates the effect of smoothing on the longitudinal CoP metric.

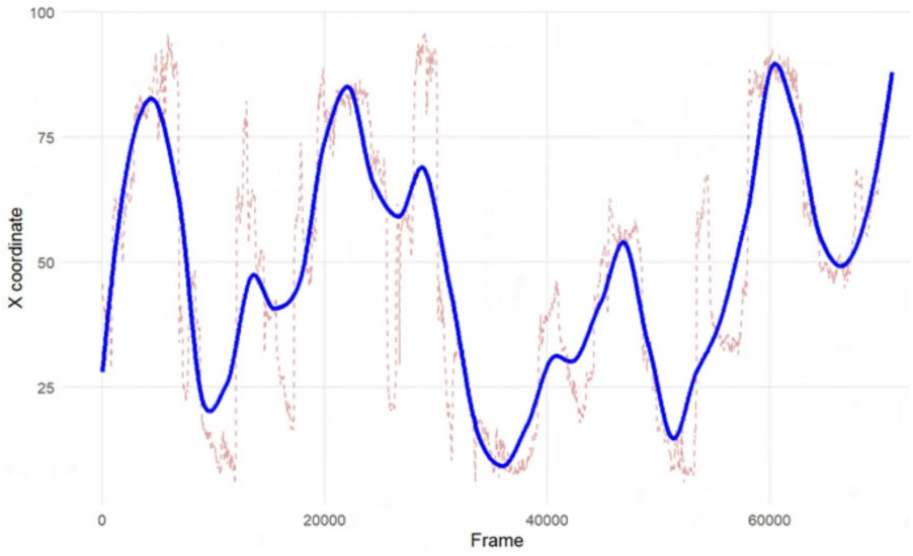


Figure 1: Smoothing of collective movement features: Centre of Play (CoP) longitudinal trajectory, raw (dashed) vs. smoothed (solid)

Local extrema of MoP. To capture critical phase transitions in team organisation, local minima and maxima were extracted from the smoothed MoP using a rolling window approach. A local minimum in MoP identifies the moment of maximum spatial concentration, consistent with the spatial compression observed during breakdown phases such as rucks or tackles. Conversely, a local maximum reflects the point of greatest formation dispersion, indicating the full deployment of an attacking line before a tactical reset. These extrema were encoded as binary indicator variables. Figure 2 demonstrates how extreme points in the smoothed MoP act as temporal markers of tactical reorganisation.

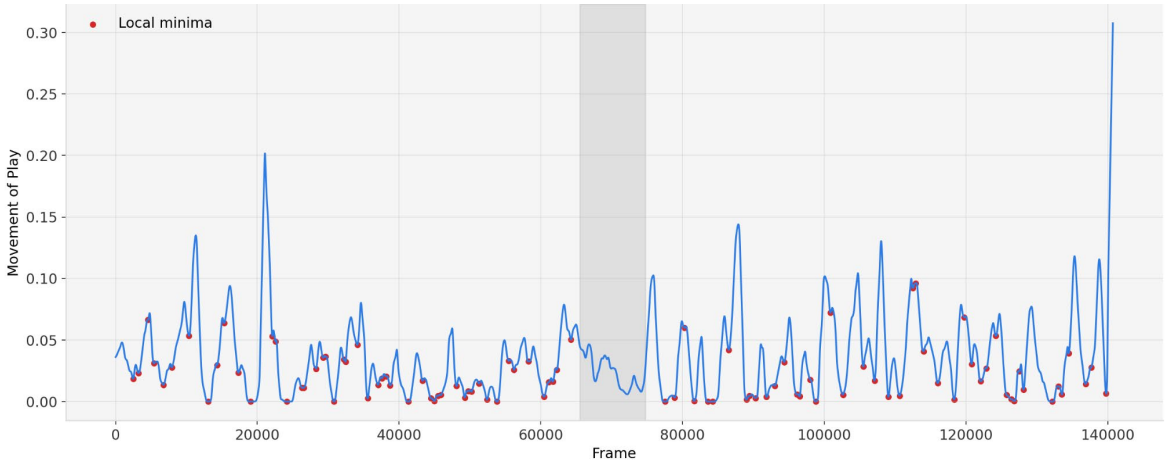


Figure 2: Local minima of the smoothed Movement of Play (MoP) trajectory, marking moments when relative collective movement is at its lowest. The shaded area is halftime.

Spatial context and event information. Each frame was assigned a categorical indicator of longitudinal pitch region, based on whether the CoP lay below or above the median longitudinal position across all frames. Additionally, restart events (including kick-off, lineout, scrum, and ruck) were manually annotated from video footage and included as a categorical predictor to provide contextual information about the phase of play. Event annotations were included as contextual covariates and are not deterministically linked to possession labels, ensuring the model does not trivially infer possession from event type alone.

4.3 Spatial Segmentation

Tactical behaviour and possession dynamics in Rugby Union vary substantially across pitch regions (Colomer et al., 2020). To account for this spatial variation, the pitch was segmented into multiple tactically meaningful regions, and possession modelling was conducted at both global and regional levels. Figure 3 provides a combined schematic of all spatial regions: the longitudinal 22–50–22 structure (defensive, neutral, and attacking zones), the lateral corridor partition (wide left, central, and wide right), and the central interaction zone (Model 7) defined as the intersection of the central longitudinal band and the central lateral corridor; the primary area where most structured play, rucks, and possession contests occur.

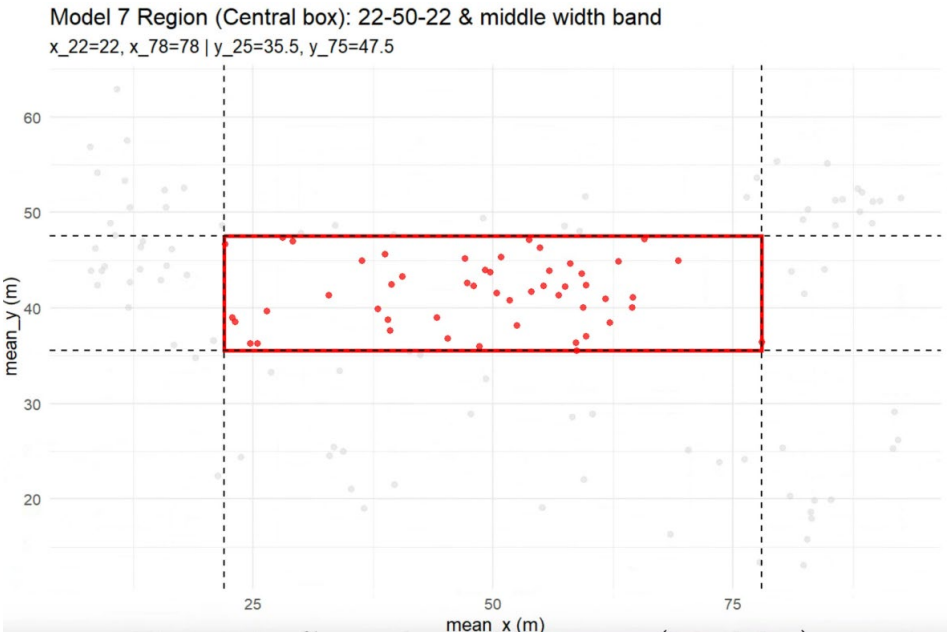


Figure 3: Combined spatial segmentation schematic: longitudinal zones (defensive / neutral / attacking, 22–50–22 structure), lateral corridors (wide left / central / wide right), and the central interaction region (Model 7, highlighted) shown on a single pitch diagram

4.4 Model Specification

Possession is modelled as a binary outcome $Y_t \in \{0, 1\}$ at time frame t relative to a designated focal team. A feature vector is constructed for each frame using the spatio-temporal descriptors above, including CoP, MoP, dispersion measures, movement dynamics, spatial region indicators, and contextual event information.

A Random Forest classifier is employed to model the complex and non-linear relationship between collective movement patterns and possession state (Breiman, 2001). Random Forests are ensemble learning methods that combine multiple decision trees trained on bootstrapped samples with random feature subsampling at each split. Two modelling strategies are considered: (i) a global possession model trained on all frames across the full pitch, and (ii) region-specific models trained on spatially defined subsets of the data. Random Forests were selected as a baseline due to interpretability and robustness; future work will explore sequence-based models such as Hidden Markov Models and LSTM networks that explicitly model temporal dependencies.

4.5 Evaluation

The dataset was partitioned into training (70%) and test (30%) subsets using a random frame-level split. Classification performance was evaluated using overall accuracy and the area under the receiver operating characteristic curve (AUC). AUC is particularly well suited to this setting, where possession durations are highly variable and class imbalance may arise across match phases (Fawcett, 2006). To assess performance stability, repeated random-split validation was performed over 30 independent partitions; mean and standard deviation of accuracy and AUC were reported. It is acknowledged that random frame-level splitting does not account for temporal autocorrelation between adjacent frames; future work will evaluate generalisation under strict temporal partitioning.

5. RESULTS

5.1 Global Model Performance

The global Random Forest model achieved an overall accuracy of 0.762 and an AUC of 0.836 (Table 2), indicating good discriminative ability in distinguishing possession status using spatio-temporal player tracking features alone. The ROC curve (Figure 5) exhibits clear separation from the diagonal reference line, confirming that the classifier substantially outperforms random guessing.

Model	Accuracy	AUC
Random Forest (Global)	0.762	0.836

Table 2: Performance of the global Random Forest possession model.

5.2 Repeated Cross-Validation

Repeated random-split validation over 30 iterations confirmed the predictive stability of the framework. Mean accuracy reached 0.796 (±0.069) and mean AUC remained consistently high at 0.867 (±0.057) (Table 3). The low standard deviations relative to the mean values suggest that the collective spatial features and tactical indicators provide a stable signal for possession identification, regardless of specific data splitting.

Metric	Mean	Standard Deviation
Accuracy	0.796	0.069
AUC	0.867	0.057

Table 3: Model performance metrics averaged over 30 random validation repeats.

5.3 Variable Importance

Variable importance analysis indicates that possession inference is most strongly associated with longitudinal territorial structure. The smoothed CoP and raw longitudinal CoP position consistently rank as the most influential predictors, reflecting sustained pitch position control by the team in possession. Measures of collective lateral dispersion contribute additional discriminative power. Contextual spatial features are substantially more informative than instantaneous movement metrics such as speed and direction. Local extrema of MoP play a minimal role in overall classification but serve as interpretable phase-transition markers.

5.4 Region-Specific Model Performance

Table 4 summarises accuracy and AUC values for all seven region-specific models, and Figure 7 shows the corresponding ROC curves. Substantial variation in predictive performance was observed across regions. Models trained in territorially structured areas generally achieved higher AUC values. The attacking 22 region achieved an AUC of 1.00 (Accuracy = 0.889), though this result should be interpreted cautiously given the small regional sample size. The central interaction region achieved an AUC of 0.964, confirming that possession signals are strongest in the primary tactical zone. In contrast, the wide left corridor exhibited weaker performance (AUC = 0.633), attributable to lower player density and greater data sparsity in wide areas. The convergence in accuracy between the global model (76.2%) and the central interaction zone (73.3%) confirms that global possession inference is predominantly driven by movement dynamics in the central pitch

.

Region	Model	Accuracy	AUC
Central interaction zone	Model 7	0.733	0.964
Attacking 22	Model A	0.889	1.000
Between 22-metre lines	Model B	0.720	0.886
Defensive 22	Model C	0.667	0.833
Wide left corridor	Model 1	0.545	0.633
Central lateral corridor	Model 2	0.762	0.897
Wide right corridor	Model 3	0.818	0.933

Table 4: Performance of region-specific Random Forest possession models.

6. DISCUSSION

6.1 Interpretation of Key Predictors

The dominance of CoP features as the most influential predictors is consistent with the territorial nature of Rugby Union. The team in possession typically maintains control over pitch position by advancing the point of play and constraining the opposition within their defensive territory. Sustained territorial occupation therefore serves as a reliable spatial indicator for the possession state, capturing longer-term patterns of territorial dominance that characterise attacking phases. The consistent superiority of the smoothed CoP over the raw value further indicates that possession is more closely associated with persistent spatial trends than with instantaneous positional fluctuations.

Collective lateral dispersion outperforms longitudinal dispersion as a predictor, suggesting that the spatial stretching of pitch width is a more sensitive indicator of possession transitions than team depth. This reflects tactical behaviours such as defensive compression and attacking width deployment during phase transitions.

6.2 Spatial Heterogeneity and Regional Modelling

The strong performance of region-specific models highlights the role of spatial context in shaping the relationship between collective movement patterns and the possession state. Team formations and movement dynamics in Rugby Union vary substantially across pitch regions. In attacking territory, teams typically expand laterally and deploy structured attacking shapes, whereas in defensive territory formations become more compact and reactive. Modelling possession separately within spatial regions allows the classifier to learn region-dependent associations between collective spatial features and possession, yielding substantially improved discrimination in tactically structured areas.

6.3 Broader Implications: Latent State Inference Beyond Possession

The results of this study carry implications beyond possession inference in Rugby Union. This framework demonstrates that meaningful tactical state information is identifiable from collective movement patterns without observing the specific event of interest. This principle extends naturally to other latent states in invasion sports, such as defensive organisation, line-break probability, and territorial pressure, and may have potential applicability in domains beyond sport where latent state inference from group movement data is of interest.

More broadly, this study contributes to a growing literature demonstrating that the collective spatial organisation of players carries sufficient information to infer possession state with meaningful reliability, opening the door to a fully automated performance analysis pipeline without requiring specialised ball-tracking infrastructure.

6.4 Accuracy, AUC, and Possession Uncertainty

Across both global and regional models, discrepancies between accuracy and AUC reflect the inherent uncertainty of possession inference in Rugby Union. High AUC values alongside more modest accuracy scores indicate that the model effectively distinguishes possession from non-possession in a probabilistic sense, but hard classification at a fixed threshold may be unreliable during highly contested and congested phases (breakdowns, rucks, and tackles) where both teams' formations overlap. The model appropriately assigns intermediate possession probabilities in these cases, reflecting genuine tactical ambiguity rather than predictive failure.

6.5 Limitations and Future Directions

Several limitations must be acknowledged. First, this study is intentionally scoped as a proof-of-concept framework. Analysis is based on a single competitive match; the focus is on demonstrating signal existence rather than universal generalisation, and future work should extend the framework to multi-match and multi-team datasets to assess robustness and transferability.

Second, the study relies on broadcast video footage where player tracking is restricted to the camera's field of view. Off-screen players may bias collective movement metrics; future research could investigate whether wide-angle tactical feeds or synchronised GPS data would refine spatial features and enhance discriminative power.

Third, model evaluation relied on random frame-level train–test splitting, which implicitly assumes frame independence. In practice, player tracking data exhibit strong temporal dependence, and adjacent-frame autocorrelation may produce optimistic performance estimates. Block-wise temporal splitting and phase-based cross-validation represent important next steps for more conservative and realistic performance evaluation.

Finally, while Random Forests provide a strong and interpretable baseline, future work should investigate sequence-based models, such as Hidden Markov Models or Long Short-Term Memory networks, that explicitly model temporal dependencies and possession transitions over longer horizons.

7. CONCLUSIONS

This study established that ball possession in Rugby Union is identifiable as a latent spatio-temporal state using player tracking data alone. The global Random Forest model achieved an accuracy of 76.2% and an AUC of 0.836, validated through 30 repeated random-split procedures. The study makes three contributions to the literature. First, it provides the first demonstration of automated possession inference in Rugby Union without ball tracking. Second, it reframes possession as an emergent property of coordinated collective movement as a latent state that is inferrable from spatial behaviour. Third, it provides empirical evidence of spatial heterogeneity in possession signals, with the central interaction zone achieving an AUC of 0.964.

Longitudinal CoP features and spatial context variables consistently proved to be the most influential predictors, reinforcing the fundamentally territorial nature of possession dynamics in Rugby Union. A combination of global and region-specific modelling offers the best balance of interpretability and discriminative power for practical deployment. More broadly, this work contributes to the emerging understanding that hidden game states are inferable from collective motion patterns, a principle that extends beyond possession in rugby to tactical state modelling across invasion sports and, potentially, to other domains where latent group-state inference from positional data is of interest.

References

Bartlett, R., Button, C., Robins, M., Dutt-Mazumder, A., & Kennedy, G. (2012). Analysing team coordination patterns from player movement trajectories in football: Methodological considerations. International Journal of Performance Analysis in Sport, 12(2), 398–424.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chambers, R. M., Gabbett, T. J., Gupta, R., Josman, C., Bown, R., Stridgeon, P., & Cole, M. H. (2019). Automatic detection of one-on-one tackles and ruck events using microtechnology in rugby union. Journal of Science and Medicine in Sport, 22(7), 827–832.

Colomer, C. M. E., Pyne, D. B., Mooney, M., McKune, A. J., & Serra-Olivares, J. (2020). Performance analysis in rugby union: A critical systematic review. Sports Medicine – Open, 6(1), Article 4. https://doi.org/10.1186/s40798-019-0232-x

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. ACM Computing Surveys, 50(2), Article 22.

Guerrero-Esteban, S., Solé, J., & Daza, G. (2023). Variables involved in ball possession in rugby: A systematic review. Apunts Educación Física y Deportes, 153, 90–100. https://doi.org/10.5672/apunts.2014-0983.es.(2023/3).153.08

Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions in elite level rugby union using a wearable sensing device. Sports Engineering, 15, 81–92. https://doi.org/10.1007/s12283-012-0088-5

Kovalchik, S. A. (2023). Player tracking data in sports. Annual Review of Statistics and Its Application, 10, 677–697. https://doi.org/10.1146/annurev-statistics-033021-110117

Link, D., & Hoernig, M. (2017). Individual ball possession in soccer. PLOS ONE, 12(7), e0179953. https://doi.org/10.1371/journal.pone.0179953

Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2020). A systematic review of collective tactical behaviours in football using positional data. Sports Medicine, 50(2), 343–385. https://doi.org/10.1007/s40279-019-01194-7

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257–286.

Vidal-Codina, F., Evans, N., El Fakir, B., & Billingham, J. (2022). Automatic event detection in football using tracking data. Sports Engineering, 25, Article 18. https://doi.org/10.1007/s12283-022-00381-6

COMPUTER VISION FOR HORSE RACING ANALYSIS AND ADJUDICATION

Anthony Bedford [a,d], Erica Mealy [a], Abigail Koay [a], Utkarsh Kumar [a], Shahrzad Saremi [a]

University of the Sunshine Coast, Queensland, Australia

Corresponding author: abedford@usc.edu.au

Abstract

Racing protests are currently adjudicated by stewards reviewing broadcast footage manually, a process that is time-intensive and inherently subjective. This paper presents a computer vision pipeline for automated horse and jockey detection, multi-object tracking, and protest adjudication support in thoroughbred racing, building on prior work (Mealy et al., 2025). We assemble a novel two-class dataset of 7,290 training images using Roboflow SAM3 automated segmentation with full manual annotation correction, a fivefold increase over our previous manually labeled baseline. We then conduct a controlled comparison of YOLOv12x (Tian et al., 2025) and RF-DETR (Robinson et al., 2025) on a held-out test set of 236 images, on which RF-DETR (ep72) achieves mAP50 of 0.920 and mAP50-95 of 0.649, outperforming YOLOv12x across 10 of 12 evaluation metrics despite training at lower resolution. We further compare three modern multi-object trackers and demonstrate that full-race identity consistency remains an unsolved problem, with our best configuration producing 1,109 unique track IDs for a 16-horse race that should yield 16. To address this in practice, we introduce a steward-centred review workflow in which the operator selects the two horses involved in a specific protest and projected proximity analysis is performed only on that pair within a short incident window. We validate this workflow against a published Racing NSW stewards inquiry (Goulburn, 2018), where the system produces a projected overlap signal beginning at t = 27.16 s, consistent with the upheld protest finding, while acknowledging that projected image-plane overlap is not equivalent to verified physical contact.

Keywords: *Computer Vision, Horse Racing, Multi-Object Tracking, Object Detection, Protest Adjudication, Artificial Intelligence*

1. Introduction

Every year, racing protests are adjudicated by stewards working from a single oblique broadcast camera feed, a process that is manual, time-intensive, and inherently subjective. When two horses make contact at the 100-metre mark, the difference between an upheld and dismissed protest rests on the judgement of an official reviewing low-resolution footage at normal playback speed, with no objective measurement of whether contact occurred or how significant it was. Computer vision systems that can quantify horse proximity from broadcast footage could fundamentally change this process, providing stewards with reproducible, timestamped evidence rather than relying solely on visual interpretation.

This challenge, however, is technically demanding. Horse racing violates nearly every assumption made by standard multi-object tracking benchmarks. Horses are visually similar in body shape, silhouette, and coat colour; they move at speeds of up to 60 km/h; they frequently bunch into packs causing sustained mutual occlusion; and footage is typically recorded from a fixed oblique camera that collapses the depth dimension, making it difficult to distinguish lateral contact from head-to-tail proximity.

This paper presents a computer vision pipeline from dataset construction through to a proof-of-concept steward review workflow, extending prior work in this domain (Mealy et al., 2025). We make three central arguments: first, that automated SAM3-based labeling with manual correction can produce a training dataset at a scale that was previously infeasible; second, that RF-DETR's transformer architecture, backed by DINOv2 pretraining, outperforms YOLOv12x for horse detection under the occlusion conditions characteristic of pack racing; and third, that full-race identity consistency need not block practical protest review support if the workflow is designed around the steward's actual task, focusing on two horses within a short incident window, while acknowledging the remaining limitations of image-plane proximity analysis.

This paper makes the following contributions:

(1) An auto-labeled dataset of 7,290 training images of racehorses and jockeys, assembled using Roboflow SAM3 automated segmentation with manual annotation correction, five times larger than our previous manually labeled baseline, and sourced from thoroughbred racing venues across Australia and New Zealand.

(2) A controlled empirical comparison of YOLOv12x and RF-DETR on both validation and held-out test sets, demonstrating that the DINOv2 backbone (Oquab et al., 2024) provides stronger feature representations even when trained at lower resolution.

(3) A tracker comparison across three modern multi-object trackers (DeepOCSort (Maggiolino et al., 2023), BoT-SORT (Aharon et al., 2022), and ByteTrack (Zhang et al., 2022) on the same detection stream, quantifying identity consistency across 300 frames of race footage.

(4) A steward-centred review workflow in which the operator selects the two horses involved in a specific protest and the system computes a frame-by-frame projected proximity timeline restricted to that pair within a short incident window.

(5) Proof-of-concept validation of the projected proximity workflow against a published Racing NSW stewards inquiry (Goulburn, 15 March 2018) (Racing NSW, 2018), with explicit acknowledgment of the limitations of image-plane proximity as a decision-support measure.

2. Related Work

Object detection in sport has been studied extensively for team sports (soccer, player detection) and individual sports (tennis, swimming), but horse racing has received comparatively little attention. Prior work in this domain has explored centroid-based speed estimation and horse detection from broadcast footage (Bedford, 2024; Mealy et al., 2025). The racing context differs from pedestrian tracking benchmarks in the high visual similarity between subjects, pack formation dynamics, and a fixed oblique camera angle that collapses depth.

General-purpose detectors have evolved rapidly. YOLOv12 introduces an Area Attention mechanism for efficient large-receptive-field modelling (Tian et al., 2025). RF-DETR adapts the DETR transformer framework with a DINOv2-pretrained backbone and neural architecture search on the encoder, becoming the first real-time detector to exceed 60 mAP on COCO (Robinson et al., 2025); DINOv2 is a self-supervised ViT providing strong transferable representations (Oquab et al., 2024).

For multi-object tracking, the SORT framework (Bewley et al., 2016) established the Kalman filter and Hungarian algorithm paradigm. ByteTrack (Zhang et al., 2022) extended this by associating low-confidence detections, while BoT-SORT (Aharon et al., 2022) added appearance-based re-identification and camera motion compensation. DeepOCSort (Maggiolino et al., 2023) integrates appearance embeddings into the OC-SORT framework. HOTA (Luiten et al., 2021) provides a unified tracking metric that balances detection and association accuracy, addressing the limitations of MOTA, which is biased toward detection, and IDF1, which is biased toward association.

SAM 3 (Carion et al., 2025) introduced a memory-based video object segmentation mechanism capable of propagating instance masks across frames. Roboflow's SAM3 extends this framework with an additional tracking pass and is the version used in this work both for auto-labelling and for segmentation tracking. Segmentation-based tracking of this family is reported to perform strongly on benchmarks characterised by heavy occlusion and visually similar subjects (Carion et al., 2025), conditions analogous to horse racing. Grounding DINO (Liu et al., 2024) enables open-set object detection from text prompts, facilitating zero-shot auto-labelling without domain-specific annotation.

3. Dataset and Auto-Labelling Pipeline

3.1 Previous Baseline

In our prior work, 1,500 frames were manually annotated with horse bounding boxes only, with no jockey class included. Manual annotation at this scale required approximately 30 person-hours and produced a single-class detector achieving 68 FPS with 0.84 average confidence on test clips.

3.2 Auto-Labelling with Roboflow SAM3 and Manual Annotation Correction

To overcome the annotation bottleneck, we first attempted zero-shot auto-labelling with Grounding DINO (Liu et al., 2024) via Autodistill; this failed on pack-racing footage due to severe mutual occlusion, producing excessive false positives and missed detections. We then used Roboflow's SAM3-based auto-annotation (Dwyer et al., 2024), which proposes bounding boxes from instance segmentation. Raw footage was sourced from venues across Australia (Melbourne, Sydney, Sunshine Coast, Queensland), New Zealand and Hong Kong, providing visual diversity across track conditions, lighting, and horse colouring. This yielded 3,138 raw annotated images.

SAM3 produced systematic errors on heavily occluded groups: where two or three horses were closely packed, it generated a single bounding box around the entire group rather than individual boxes. Every frame was therefore manually reviewed in Roboflow's labelling interface, and merged boxes were split into the correct number of individual horse and jockey annotations.

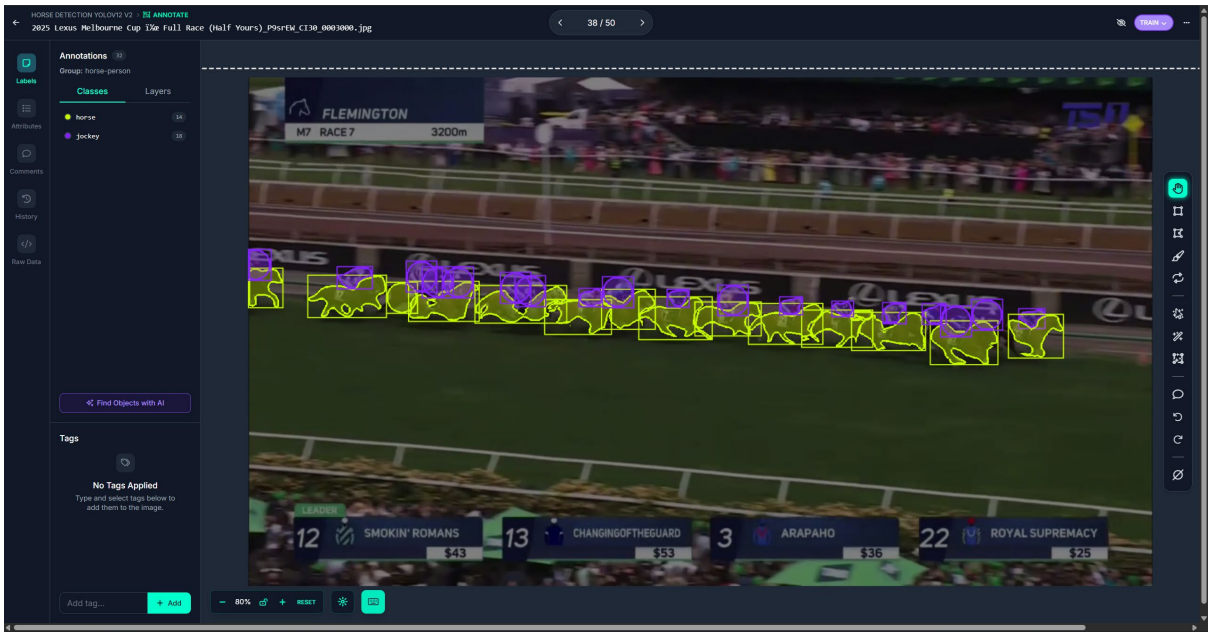


Figure 1: Automatic horse annotation using Roboflow's SAM 3 auto-labeling tool

From the 3,138 raw images, 2,430 training images were augmented three-fold using horizontal flip, brightness adjustment of +/-15%, exposure adjustment of +/-10%, Gaussian blur up to 1.4 px, and additive noise up to a factor of 1.05x, yielding 7,290 training images. The validation set (472 images) and test set (236 images) were not augmented, preserving their role as unmodified held-out evaluation sets. The final dataset totals 7,982 annotated images split 91.4% / 5.9% / 2.9% across train, validation, and test, provided in both YOLO and COCO formats for cross-model compatibility.

3.3 Evaluation Clips

Table 1. Video clips used for system evaluation.

Clip	Frames	Duration	Resolution	Horses	Primary use
7-horse clip	875	35 s	640x360	7	Detection eval; tracking comparison
6-horse clip	806	26.9 s	1920x1080	6	Operational eval (primary)
Full-race clip	4,624	154 s	1920x1080	~16	ID fragmentation quantification

Three video clips were used for operational and tracking evaluation (Table 1). The test split of the static dataset was held out from all training and hyperparameter selection decisions, and no evaluation clip was used during training.

4. Object Detection: YOLOv12x versus RF-DETR

4.1 Model Architectures

YOLOv12x is the extra-large variant of the YOLOv12 series. It introduces Area Attention (A2), an attention mechanism with linear complexity that provides large receptive fields, and R-ELAN modules for stable gradient flow in deep attention networks (Tian et al., 2025).

RF-DETR (large variant) uses a DINOv2-pretrained ViT-L backbone connected to a lightweight deformable DETR decoder (Robinson et al., 2025; Zhu et al., 2021). The global self-attention in the ViT backbone models long-range spatial dependencies across the full input image rather than relying solely on local convolutional receptive fields. This property is directly relevant to pack-racing occlusion, in which a partially hidden horse must be localised using contextual cues from across the frame.

4.2 Training Setup

Both models were trained on the same dataset using a cloud GPU (NVIDIA RTX 5090, 32 GB VRAM). RF-DETR was trained at resolution 640 with a batch size of 16 and a gradient accumulation factor of 2 (effective batch size of 32), for 72 epochs with early stopping at patience 20, using rfdetr library v1.6.2 and a confidence threshold of 0.35. YOLOv12x was trained at imgsz 832 with batch size 32 for 58 epochs with early stopping, using Ultralytics v8.3.204, with per-class confidence thresholds of 0.40 for horse and 0.23 for jockey optimised on the validation set. Both models were evaluated on the same held-out test set of 236 images using the standard COCO evaluation protocol.

4.3 Detection Results on the Held-Out Test Set

Table 2. Per-class detection results on the held-out test set (236 images). Winning values are in bold. RF-DETR wins 10 of 12 metrics.

Metric	RF-DETR ep72	YOLOv12x	Winner
mAP50	**0.9196**	0.8761	RF-DETR
mAP50-95	**0.6494**	0.6358	RF-DETR
Precision	**0.9171**	0.8570	RF-DETR
Recall	0.7522	**0.8162**	YOLOv12x
Horse mAP50	**0.9206**	0.8457	RF-DETR
Horse mAP50-95	**0.6985**	0.6274	RF-DETR
Jockey mAP50	**0.9185**	0.9064	RF-DETR
Jockey mAP50-95	0.6004	**0.6443**	YOLOv12x

RF-DETR wins on 10 of 12 metrics (Table 2). Notably, this performance is achieved despite training at a lower image resolution of 640 pixels compared to 832 pixels for YOLOv12x, suggesting that the DINOv2 pretrained backbone provides stronger feature representations than resolution alone can compensate for. YOLOv12x

achieves higher recall and higher jockey mAP50-95, consistent with its higher training resolution enabling finer localisation of small jockey figures.

4.4 Operational Performance on the 6-Horse Clip

To complement formal test set evaluation, we measured operational performance on the 6-horse clip (806 frames) at recommended per-model confidence thresholds. Missed frames are those in which the detected horse count falls below the known ground truth of 6; over-detected frames are those with more detections than expected.

Table 3. Operational detection performance on the 6-horse clip (806 frames).

Model	Threshold	FPS	Missed frames	Over-detected	Avg conf
RF-DETR ep72	0.35	8.9	106	126	0.894
YOLOv12x (per-class)	h=0.40 j=0.23	11.6	215	80	0.848
YOLOv12 Sem1 baseline	0.23	32.0	458	0	0.878

RF-DETR achieves fewer missed frames at the cost of higher over-detection. YOLOv12x produces fewer false positives but misses more horses, particularly during pack occlusion. Both models substantially outperform the Semester 1 baseline, which was a horse-only single-class model trained on 1,500 images.

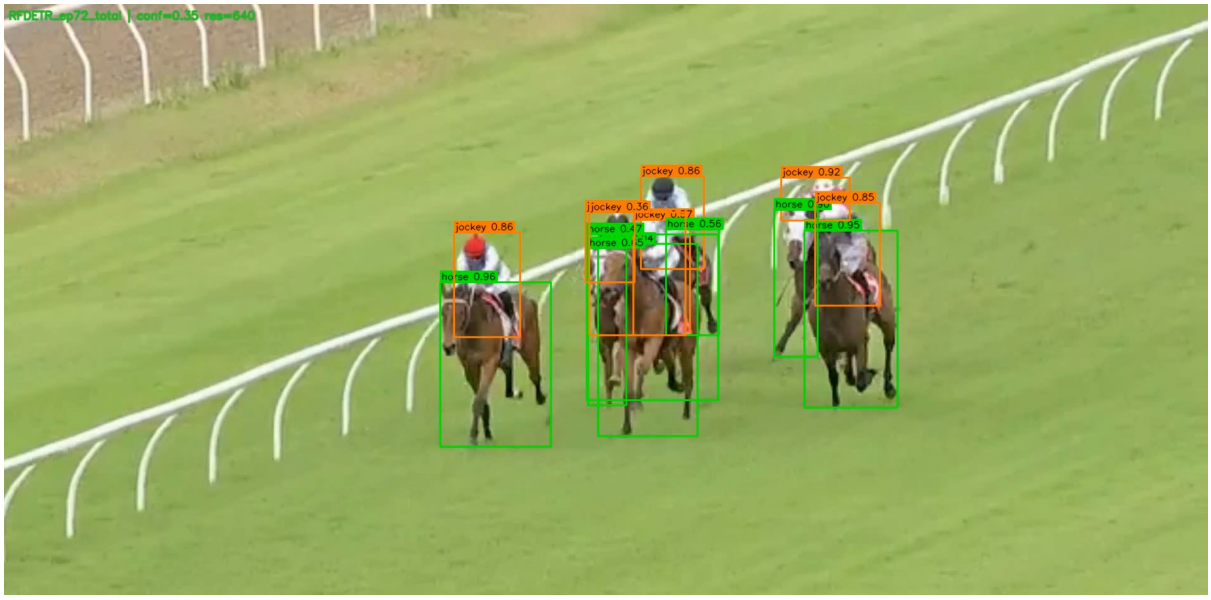


Figure 2: Horse and jockey detection results using RF-DETR, showing bounding box predictions with associated confidence scores



Figure 3: Horse and jockey detection results using YOLO, illustrating bounding box predictions and confidence scores in a racing scenario

5. Multi-Object Tracking

5.1 The Identity Consistency Challenge

Maintaining consistent identities across all horses for the full duration of a race is the central unsolved problem in this domain. Horses frequently enter tight packs in which partial occlusion is sustained for one to several seconds. When a horse re-emerges from occlusion, standard trackers typically assign a new identity rather than recognise the previously tracked animal, producing fragmented track histories.

To quantify the problem at race scale, we ran both RF-DETR and YOLOv12x with DeepOCSort tracking on a complete 16-horse race clip (4,624 frames, 154 seconds). RF-DETR produced 1,109 unique track IDs and YOLOv12x produced 1,783 unique track IDs for what should theoretically be 16 tracks. These results demonstrate that identity fragmentation becomes qualitatively worse at race scale and cannot be resolved with current trackers alone.

5.2 Tracker Comparison

To compare modern trackers fairly, RF-DETR detections were passed through three trackers simultaneously on the same 300-frame segment of the 6-horse clip: DeepOCSort (Maggiolino et al., 2023), BoT-SORT (Aharon et al., 2022), and ByteTrack (Zhang et al., 2022). All trackers processed identical detections; differences in output reflect only the tracking algorithm. Fewer unique horse IDs assigned over the same footage implies fewer fragmentation events; for 6 horses over 300 frames, the theoretical minimum is 6 unique IDs.

Table 4. Unique horse track IDs assigned over 300 frames of the 6-horse clip (RF-DETR detections). Theoretical minimum: 6. Fewer is better.

Tracker	Unique horse track IDs	Notes
DeepOCSort	15	OC-SORT with appearance embeddings
BoT-SORT	25	Kalman filter + ReID + camera motion compensation
ByteTrack	31	Associates all low-confidence detections

DeepOCSort assigns only 15 unique horse track IDs across 300 frames, compared to 31 for ByteTrack and 25 for BoT-SORT (Table 4), indicating stronger re-association after occlusion. A custom horse re-identification model was also developed using a ResNet-50 backbone trained on horse appearance crops extracted from race footage. Despite domain-specific training, the custom model performed below the tracker's default OSNet appearance embedding and was not adopted for final evaluation.

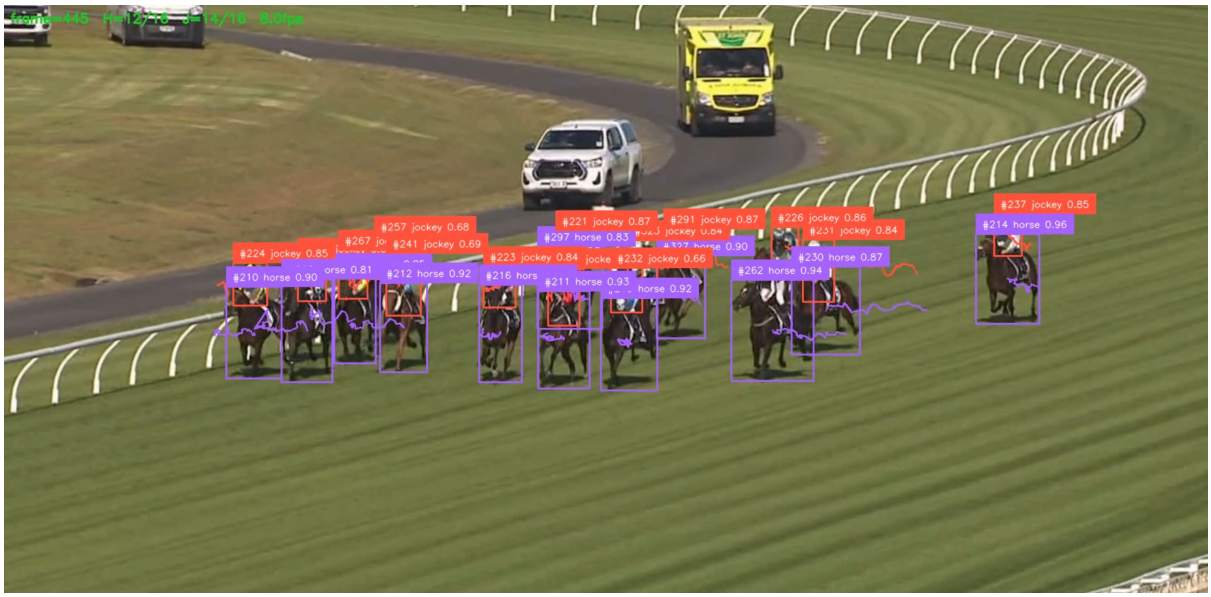


Figure 4: Multi-object tracking of horses and jockeys using DeepOCSORT, showing persistent IDs and motion trajectories across frames

5.3 SAM3 Segmentation Tracking

We evaluated SAM3 (Carion et al., 2025) as an alternative tracking approach, using RF-DETR detections from the first 30 frames as bootstrap prompts and SAM3's memory mechanism to propagate masks through subsequent footage. On the runs reported here, SAM3 achieved zero identity switches on both the 7-horse clip (7 unique tracks, 875 frames, average confidence 0.963) and the 6-horse clip (6 unique tracks, 806 frames, average confidence 0.971) on a cloud GPU. However, these clean runs were not reliably reproducible: on repeated runs over the same 6- and 7-horse clips, SAM3 intermittently exhibited the same identity-collapse failure mode described below, in which only one horse was segmented and every track ID was assigned to that single animal.

The zero-switch figures above therefore reflect successful runs rather than consistent behaviour on small-clip footage.

On the full-race clip (4,624 frames, approximately 16 horses), this failure mode was consistent rather than intermittent: all track IDs progressively converged onto a single horse's mask, with SAM3's memory mechanism propagating a single dominant identity across all subjects rather than maintaining separate tracks. This failure mode, distinct from the annotation merging issue encountered during auto-labeling, caused the system to lose all but one track by the midpoint of the clip. SAM3 also exhausted cloud GPU memory on this clip, indicating that per-object memory requirements scale super-linearly with subject count. These resource and correctness constraints make SAM3 impractical for routine operational deployment at full race scale. However, the results on



shorter clips confirm that segmentation-based tracking can achieve zero identity switches on at least some runs under controlled conditions, though not reliably, and motivate future investigation of memory-efficient segmentation trackers with more stable identity maintenance (Carion et al., 2025).

Figure 5:Failure case of SAM3-based tracking showing identity collapse, where the memory mechanism propagates a single dominant mask and assigns multiple track IDs to one horse instead of maintaining separate identities.

6. Steward Review Workflow

6.1 Motivation and Workflow

The identity consistency problem in Section 5 limits fully automated whole-race proximity analysis: if a tracker switches a horse's identity mid-race, subsequent proximity calculations may compare different animals. Rather than solving full-race identity, this study adopts a steward-centred workflow that restricts reliance on long-horizon track stability.

In a steward's inquiry the incident is already known from the protest, so the review task is focused on a short section of footage. The operator loads the video and tracking output, pauses at the alleged incident, and selects the two horses from the image; the system defines a window of roughly five seconds either side and computes a

frame-by-frame projected proximity timeline for that pair only. This mirrors how officials already review incidents and reduces — but does not eliminate — the effect of race-wide identity instability, since local switches within the selected interval remain possible. The interface is therefore a decision-support tool for incident review, not an automated adjudicator.

6.2 Proximity Metric

Earlier sports-tracking work often relied on centroid-to-centroid Euclidean distance. For racing interference this is poorly suited: centroids may remain separated even when the visible boxes overlap, making the measure difficult to interpret as close interaction.

The implemented method instead uses a signed image-plane edge-gap between the selected detections, where positive values indicate projected separation and negative values indicate projected overlap. This is more interpretable than centroid distance but is not proof of physical contact: in monocular footage, 2D overlap may arise from loose boxes, partial occlusion, or perspective compression. The metric is best understood as a projected proximity measure.

Several refinements improve robustness: direction estimation uses the bottom-centre of the detection (more stable under jockey movement than the centroid); the lateral calculation is applied to a trimmed body-contact band rather than the full box; and a temporal persistence filter suppresses isolated single-frame events.

6.3 Lateral Decomposition for Camera-Angle Robustness

A major limitation of image-plane proximity in racing footage is camera perspective: in a fixed oblique view, horses running head-to-tail can appear to overlap in 2D despite being laterally separated. Because racing interference is primarily lateral encroachment rather than longitudinal following distance, the system decomposes proximity along the estimated lateral direction of motion.

Within the analysis window, direction of travel is estimated from the median per-frame displacement of the selected horses (using bottom-centre anchoring), and a lateral axis is inferred perpendicular to that motion. Proximity is evaluated along this lateral direction rather than unrestricted 2D overlap. A specific rule is applied to head-on views, where motion is predominantly vertical, to avoid rotating a noisy motion estimate into an implausible axis.

Lateral decomposition reduces some perspective-induced false positives but is an incomplete correction: false overlaps may still occur under tight packing, loose boxes, depth compression, or local identity switches. Complete geometric correction requires homographic transformation to a calibrated bird's-eye view — beyond the present scope and identified as future work, alongside mask-based proximity using segmented body regions to better approximate side-to-side interaction.

7. Validation: Goulburn 2018 Case Study

7.1 Race Details

To validate the proximity detection capability against a known outcome, we obtained publicly available stewards footage from Race 6, Goulburn Racecourse, 15 March 2018 (Racing NSW official stewards review recording)

(Racing NSW, 2018). A protest was lodged by the connections of Fallada (barrier 3) alleging interference from Read All About It in the straight. The Racing NSW stewards upheld the protest, finding that Read All About It shifted ground and made contact with Fallada at approximately the 100-metre mark. The footage is a split-screen recording: the top half shows the broadcast oblique camera angle and the bottom half shows a head-on close-up camera aligned with the direction of travel. The video is 32.6 seconds in duration at 25 FPS and 640 x 360 resolution.

7.2 System Results

The Goulburn footage was processed through the detection and tracking pipeline to produce a frame-level tracking output for use in the steward review workflow. Because the video contains simultaneous split-screen views, the tracker produced multiple identities across both image regions, increasing the total number of unique track IDs and requiring manual identification of the relevant pair in the broadcast portion of the frame. After the relevant horses were selected in the interface, the proximity module generated a frame-by-frame projected proximity timeline over the local incident window, as shown in Table 5.

Table 5. Projected proximity timeline for the selected pair in the Goulburn 2018 stewards inquiry. Negative gap values indicate image-plane projected overlap, not verified physical contact

Event	Time	Lateral gap
First proximity flag	t = 26.84 s	+5.6 px
First projected overlap frame	t = 27.16 s	-0.7 px
Peak projected overlap	t = 27.88 s	-25.4 px
Last projected overlap frame	t = 27.96 s	n/a
Total projected-overlap duration	21 frames	0.84 s

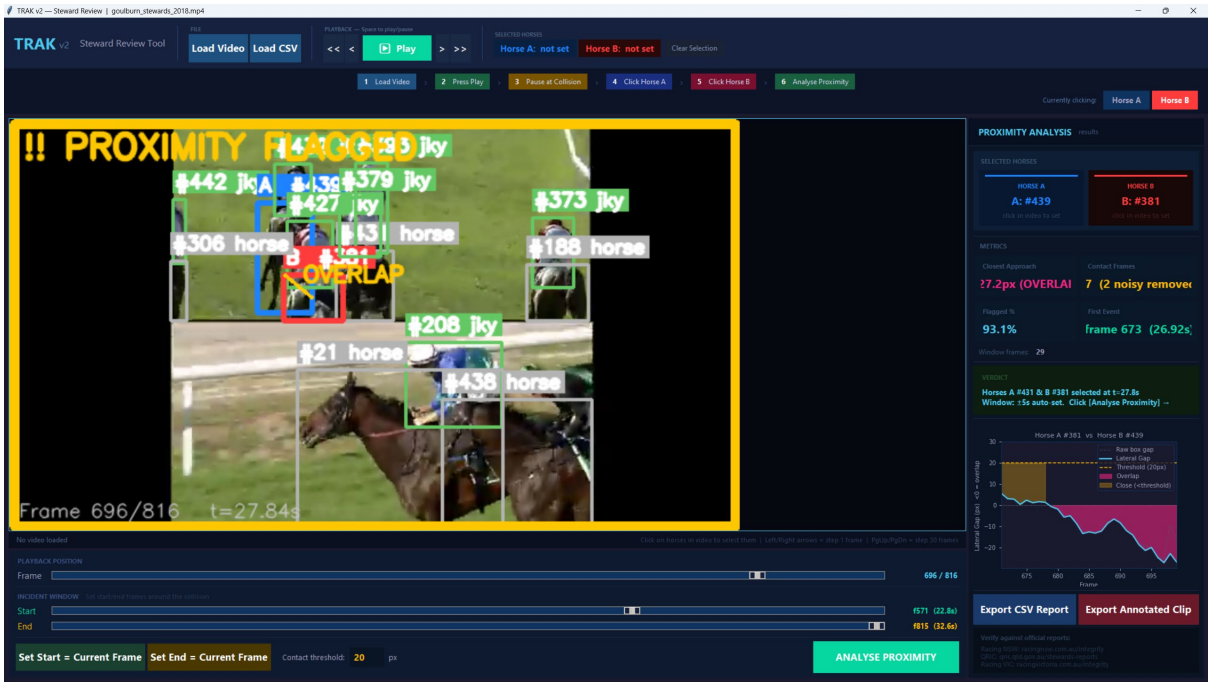


Figure 6: Steward Review Tool showing a raw box-gap false positive at t = 27.84 s. Horses #381 and #439 are longitudinally offset (front-to-back) rather than laterally adjacent, yet the 2D bounding boxes overlap due to camera projection angle. The lateral-gap co

7.3 Discussion

The Goulburn case study provides proof-of-concept evidence that the proposed steward-review workflow can localise an officially recognised incident and produce a projected proximity signal in the relevant portion of the race. The resulting timeline correctly directed attention to the incident window without requiring exhaustive manual frame-by-frame search, supporting the usefulness of the approach as a practical review aid for narrowing the search space and presenting a structured timeline of close interactions between a selected pair.

The case study also highlights limitations. Negative gap values indicate projected overlap of the detection representation, not verified physical contact, and residual false positives may occur from perspective compression, occlusion, loose boxes, or local identity instability. The split-screen structure of the Goulburn footage further inflated the identity count and required manual pair selection. The system is therefore best framed as a decision-support proximity screener.

This validation is a single known case and provides proof-of-concept rather than statistical validation. Future work should evaluate against a larger corpus of upheld and dismissed protests, investigate local pair re-association, explore mask-based proximity, and apply homographic transformation to a calibrated bird's-eye view.

8. Conclusion

This paper presented a computer vision pipeline for racehorse detection, multi-object tracking, and steward-centred protest adjudication (Mealy et al., 2025). Auto-labeling with Roboflow SAM3 and manual correction produced a dataset five times larger than the manual baseline (7,290 vs. 1,500 images) and enabled two-class horse-and-jockey detection absent from prior work. RF-DETR (ep72, DINOv2 backbone) outperformed

YOLOv12x on 10 of 12 metrics despite lower training resolution (640 vs. 832 px), supporting the suitability of transformer-based global attention under pack-racing occlusion. DeepOCSort (Maggiolino et al., 2023) achieved the fewest unique IDs (15 vs. 25 for BoT-SORT and 31 for ByteTrack) over 300 frames; SAM3 achieved zero ID switches on shorter clips but exhausted 32 GB VRAM on the full-race clip, limiting operational use.

The steward-centred workflow reduces reliance on full-race identity consistency by restricting proximity analysis to two operator-selected horses within a short window. The signed edge-gap metric with lateral decomposition and temporal filtering is more interpretable than centroid distance, though projected image-plane overlap is not verified physical contact. On the Goulburn 2018 inquiry, the system produced a projected overlap signal from t = 27.16 s to t = 27.96 s, consistent with the stewards' finding of interference at the 100-metre mark. Future work includes evaluation across a larger protest corpus, mask-based proximity, homographic correction to a bird's-eye view, and multi-camera extension.

References

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. arXiv. https://doi.org/10.48550/arXiv.2206.14651

Bedford, A. (2024). Computer vision applications in horse racing. In Proceedings of MathSport Asia 2024. University of the Sunshine Coast.

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. In Proceedings of the IEEE International Conference on Image Processing (pp. 3464-3468). https://doi.org/10.1109/ICIP.2016.7533003

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K. V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., … Feichtenhofer, C. (2025). SAM 3: Segment Anything with Concepts. arXiv. https://doi.org/10.48550/arXiv.2511.16719

Dwyer, B., Nelson, J., & Solawetz, J. (2024). Roboflow: Give your models the gift of vision [Software]. Roboflow. https://roboflow.com

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. In Proceedings of the European Conference on Computer Vision (pp. 38-55). https://doi.org/10.1007/978-3-031-72970-6_3

Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., & Leibe, B. (2021). HOTA: A higher order metric for evaluating multi-object tracking. International Journal of Computer Vision, 129(2), 548-578. https://doi.org/10.1007/s11263-020-01375-2

Maggiolino, G., Ahmad, A., Cao, J., & Kitani, K. (2023). Deep OC-SORT: Multi-pedestrian tracking by adaptive re-identification. In Proceedings of the IEEE International Conference on Image Processing (pp. 3025-3029). https://doi.org/10.1109/ICIP49359.2023.10222576

Mealy, E., Bedford, A., & Koay, A. (2025). Computer vision for horse racing analysis and adjudication. In Proceedings of the 11th MathSport International Conference. Luxembourg.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., & El-Nouby, A. (2024). DINOv2: Learning robust visual features without supervision. Transactions on Machine Learning Research. https://doi.org/10.48550/arXiv.2304.07193

Racing NSW. (2018). Stewards inquiry: Race 6, Goulburn, 15 March 2018 [Official stewards report]. https://www.racingnsw.com.au/integrity/stewards/stewards-reports/

Robinson, I., Robicheaux, P., Popov, M., Ramanan, D., & Peri, N. (2025). RF-DETR: Real-time detection transformer [Software]. Roboflow. https://github.com/roboflow/rf-detr

Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. arXiv. https://doi.org/10.48550/arXiv.2502.12524

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In Proceedings of the European Conference on Computer Vision (pp. 1-21). https://doi.org/10.1007/978-3-031-20047-2_1

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2021). Deformable DETR: Deformable transformers for end-to-end object detection. In Proceedings of the International Conference on Learning Representations. https://openreview.net/forum?id=gZ9hCDWe6ke

Mapping Artificial Intelligence and Machine Learning Research in Sports Injury Prediction: A Bibliometric Analysis

Kamalpour, Mostafa [a,b], Shibl, Rania [a,c], Saremi, Shahrzad [a], Mirzaei, Mansoureh [a], Bedford, Anthony [a], Mealy, Erica [a]

[a] *School of Science, Technology and Engineering, University of the Sunshine Coast, Brisbane, QLD 4052, Australia*
[b] *School of Information Systems, Queensland University of Technology, Brisbane, QLD 4000, Australia*
[c] *Faculty of Science and Engineering, Southern Cross University, Gold Coast, QLD, 4425, Australia*
[d] *Corresponding author: mkamalpour@usc.edu.au*

Abstract

Recent advances in artificial intelligence (AI) and machine learning (ML) have significantly enhanced sports injury prediction and prevention. However, the rapid growth of this field has led to a fragmented understanding of its intellectual structure, collaboration patterns, and emerging trends. This study addresses this gap through a bibliometric analysis of 3,066 publications from the Web of Science Core Collection (1990–2026), using VOSviewer to examine publication trends, leading contributors, collaboration networks, and thematic development. The findings indicate rapid expansion since 2015, increasing global collaboration, and a concentration of research output among leading countries, including the US, China, and the UK. Thematic analysis reveals a shift toward generalisable, data-driven models, with machine learning as the central methodological driver. This study contributes by providing a structured mapping of the field and informing future research and policy development.

Keywords: Artificial intelligence, machine learning, sports injury, bibliometrics

1. INTRODUCTION

Sports injuries remain a critical concern in both professional and amateur athletics, with significant implications for athlete performance, long-term health, and economic costs associated with treatment and rehabilitation (Maffulli et al., 2011). Traditional injury prevention approaches have largely relied on observational analysis and retrospective assessment, which often lack predictive capability in increasingly complex and data-rich sporting environments(Kalkhoven et al., 2021).

Recent advances in artificial intelligence (AI) and machine learning (ML) have introduced new opportunities to address these limitations. These techniques enable the analysis of high-dimensional datasets derived from wearable sensors, biomechanical systems, and physiological monitoring, allowing for more accurate prediction of injury risk and optimisation of training load (Bunker & Thabtah, 2019; Van Eetvelde et al., 2021). As a result, AI-driven injury prediction has emerged as a rapidly growing research domain at the intersection of sports science, data analytics, and health informatics. Despite this growth, the field remains fragmented across disciplines, making it difficult to obtain a consolidated understanding of its intellectual structure, dominant research themes, and emerging trends. Furthermore, the rapid increase in publications creates challenges for identifying influential contributors, collaboration patterns, and research frontiers.

Bibliometric analysis provides a systematic and scalable approach to addressing these challenges by mapping scientific knowledge, identifying collaboration networks, and detecting thematic structures across large datasets (Zupic & Čater, 2015). Bibliometric methods enable the analysis of knowledge diffusion and research evolution at a macro level, offering insights into both academic influence and structural development of research domains (Kamalpour et al., 2025). Therefore, to address the fragmentation of the field and provide a structured understanding of its development, this study performs a bibliometric analysis by investigating the following research questions: **RQ1:** How has AI-driven sports injury research evolved over time in terms of publication growth? **RQ2:** Which countries, institutions, and authors are most productive and how do they collaborate? **RQ3:** What are the dominant thematic clusters shaping the knowledge base of this field? **RQ4:** What emerging topics and research trends can be identified through keyword evolution over time? By addressing these questions, this study provides a comprehensive overview of the intellectual, collaborative, and thematic landscape of AI-driven sports injury research.

2. METHODS

This study adopts a bibliometric approach to systematically map the research landscape of AI and ML applications in sports injury prediction and prevention. Bibliometric methods are widely used for analysing large-scale scientific datasets, enabling the identification of publication trends, collaboration networks, and thematic structures across research domains (Zupic & Čater, 2015). Bibliometric analysis provides a scalable

and structured means of examining knowledge diffusion and intellectual development within a field, particularly when dealing with interdisciplinary and rapidly evolving topics (Kamalpour et al., 2025).
The Web of Science (WoS) Core Collection was selected as the primary data source due to its rigorous indexing standards and extensive coverage of high-quality peer-reviewed publications (Moed, 2005; Mongeon & Paul-Hus, 2016). A comprehensive search strategy was developed to capture the intersection of three domains: (1) AI and ML techniques, (2) injury-related concepts, and (3) sports-related contexts. The search was applied to the "*Topic"* field in WoS and constructed as follows:
AI/ML terms: "machine learning" OR "artificial intelligence" OR "deep learning" OR "neural networks" OR "random forest" OR "support vector machine" OR "decision tree" OR "predictive model" OR "data-driven model" OR "k-means clustering" OR "hierarchical clustering" OR "principal component analysis" OR "linear regression" OR "logistic regression" OR "naive bayes classifier" OR "support vector regression" OR "feature selection" OR "hyperparameter tuning" OR "cross validation" OR "model evaluation metrics" AND
Injury-related terms: "injury prediction" OR "injury risk" OR "injury prevention" OR "sports injury" OR "musculoskeletal injury" OR "overuse injury" OR "injury surveillance" OR "injury assessment" OR "injury diagnosis" OR "risk assessment" OR "damage mitigation" OR "trauma prevention" OR "safety measures" OR "rehabilitation" AND **Sports-related terms:** "sport" OR "athlete" OR "football" OR "soccer" OR "basketball" OR "running" OR "training load" OR "workload" OR "biomechanics" OR "wearable sensors" OR "motion capture" OR "performance monitoring" OR "fatigue monitoring" OR "athletic performance" OR "exercise science" OR "physical fitness" OR "wearable technology" OR "sensor technology" OR "kinesiology" OR "performance analytics" OR "physiological monitoring".

The search was limited to publications between 1 January 1990 and 22 March 2026, resulting in 3,066 records. Only journal articles, conference papers, and review articles were included, while non-scholarly outputs were excluded. Following data retrieval, a cleaning and preprocessing stage was conducted, including standardising author names, consolidating institutional affiliations, and merging synonymous keywords (e.g., "machine learning" and "ML") to reduce fragmentation. Such steps are essential to avoid inconsistencies that may distort bibliometric networks (Van Eck & Waltman, 2010).
The analysis was conducted using VOSviewer (Van Eck & Waltman, 2010) and included four components: (1) publication trend analysis, (2) co-authorship analysis at country, institution, and author levels, (3) keyword co-occurrence analysis, and (4) overlay visualisation to identify emerging trends. Citation counts were analysed using raw totals and average citations per publication to account for publication age. Association strength normalisation was applied in VOSviewer to ensure comparability across co-occurrence data. Together, these methods provide a comprehensive mapping of the field's intellectual, collaborative, and thematic structure.

3. RESULTS

3.1 Evolution of Research

The temporal analysis of publications demonstrates a clear and rapid growth trajectory in AI-driven sports injury research. As shown in Figure 1, the number of publications remained relatively low prior to 2010, followed by steady growth, and then a sharp increase after 2015. Notably, annual publications increased from fewer than 100 papers in 2017 to over 600 papers by 2026, indicating an exponential growth phase. This trend reflects the increasing availability of sports performance data, the adoption of wearable technologies, and the growing application of machine learning techniques in sports science.

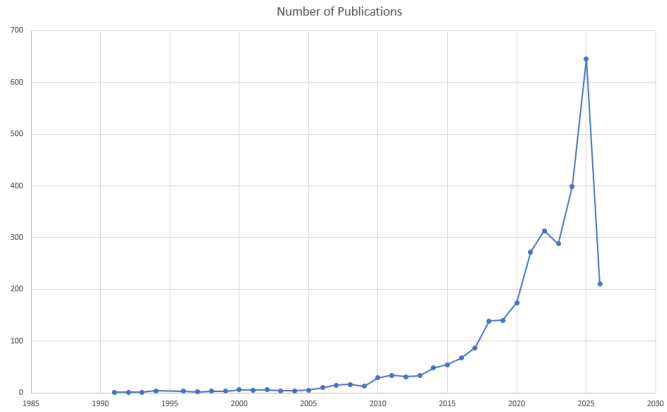


Figure 1) Number of publications per year until March 2026

3.2 Leading Contributors and Collaboration

As illustrated in Figure 2, the United States, China, and the United Kingdom emerge as the most productive countries in this research domain, with the US leading by a substantial margin (891 publications), followed by China (491) and the UK (370). Collectively, these three countries account for %57.14 of the total global research output, highlighting a strong concentration of knowledge production within a relatively small number of leading nations. This concentration suggests that research capacity, funding availability, and technological infrastructure play a significant role in shaping contributions to this field.

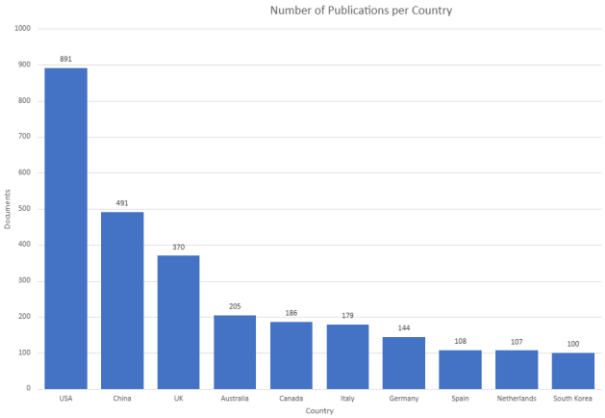


Figure 2) Top 10 countries in publication

Furthermore, as shown in Figure 3, the temporal distribution of publications reveals a notable shift in research activity over time. Earlier contributions were predominantly led by the United States, Australia, and several European countries, reflecting their early investment in sports science and data-driven research. However, more recent years show a rapid increase in contributions from China and other emerging countries, indicating a dynamic change in the global research landscape. This trend suggests not only a redistribution of research activity but also a potential shift in global research leadership, as emerging contributors increasingly influence the direction and development of AI-driven sports injury research.

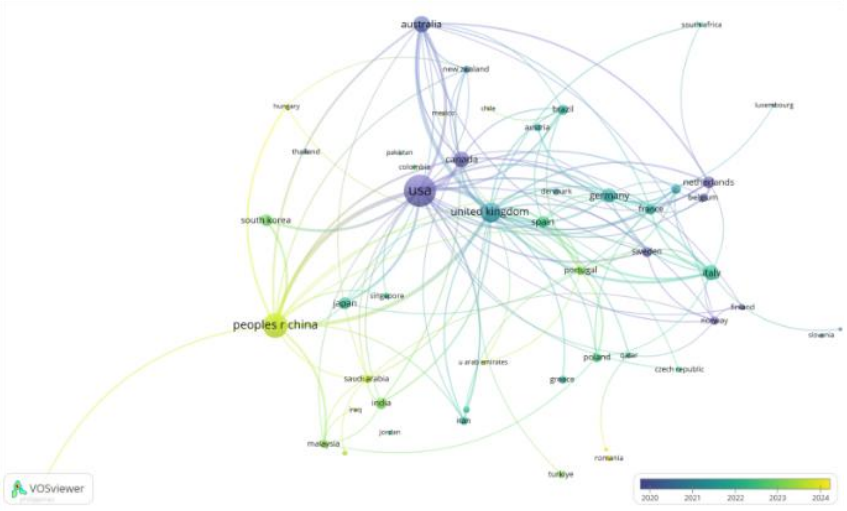


Figure 3) Co-authorship vs countries over time

The collaboration network visualisation (Figure 4) highlights strong international partnerships within this domain. The identified clusters indicate close collaboration between countries such as the US, Canada, South Korea, and China, as well as between Australia and the UK, forming key hubs within the global research network. These patterns suggest that research in this field is supported by well-established international connections, with leading countries actively collaborating and contributing to the development of the domain.

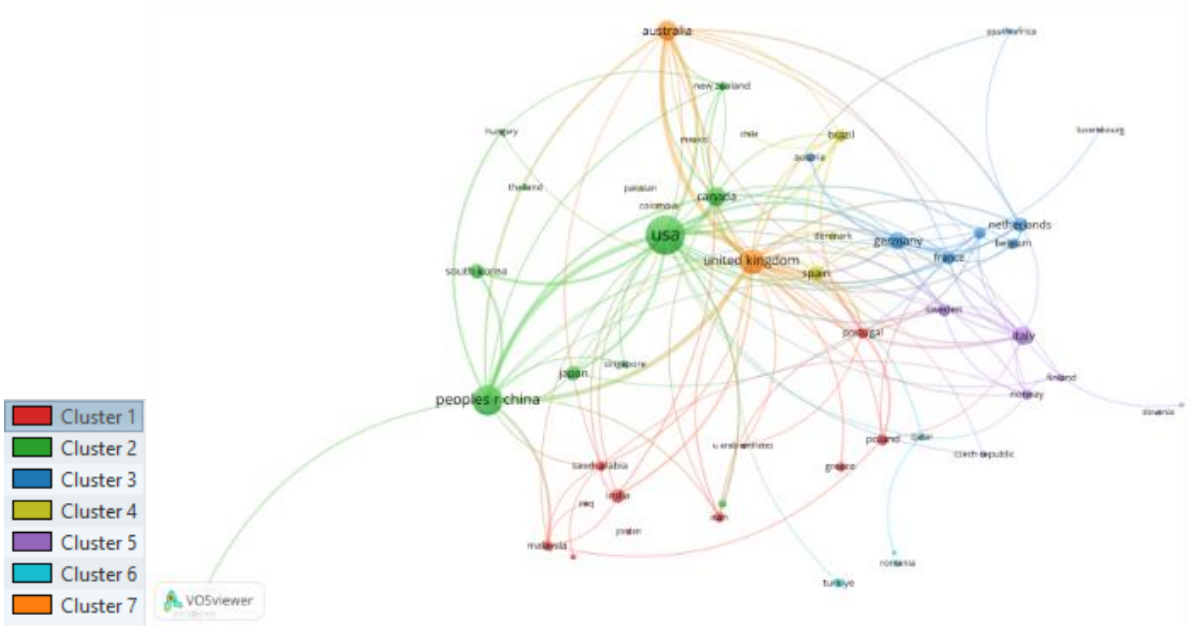


Figure 4) Country collaboration network

Institutional Contributions

Institutional analysis reveals that a relatively small number of universities dominate research output in this domain. As shown in Figure 5, leading institutions contribute a disproportionately high number of publications, indicating a strong concentration of expertise within a limited group of organisations. The collaboration network (Figure 6) further demonstrates clustered structures, where universities form closely connected research groups, reflecting established partnerships and shared research agendas. In addition, the overlay visualisation (Figure 7) suggests that institutional collaboration has intensified in recent years, particularly among Chinese universities, highlighting increased research activity and growing investment in this field.

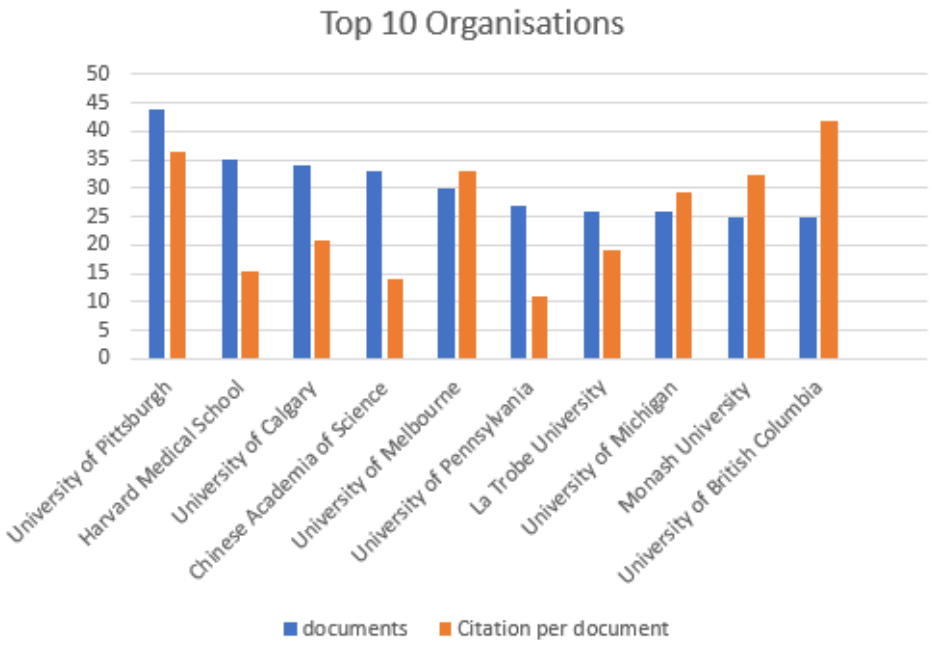


Figure 5) Top 10 institutions number of documents and citations

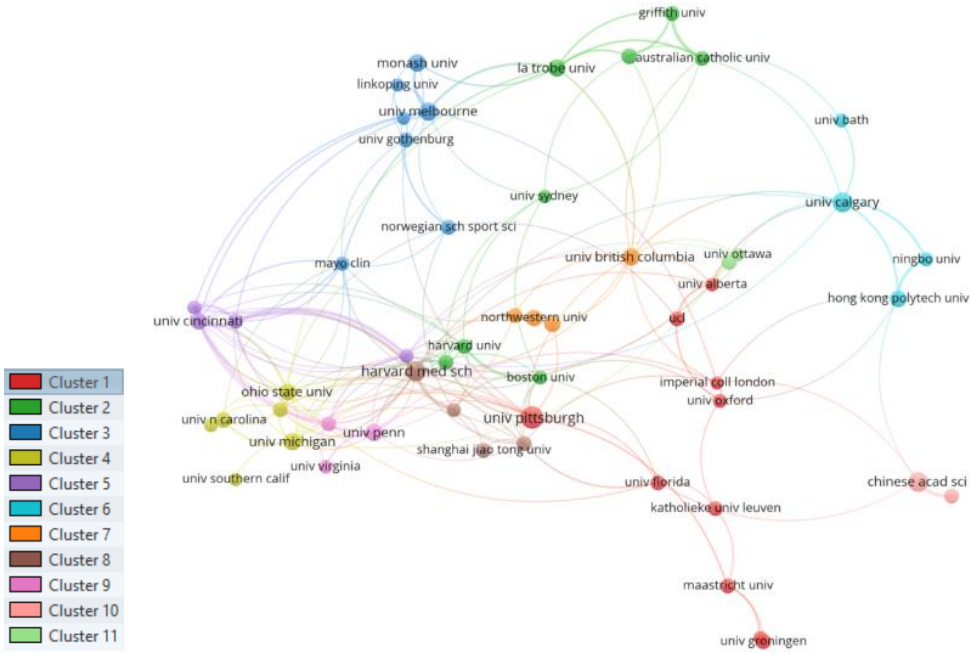


Figure 6) Clusters of collaboration among organisations

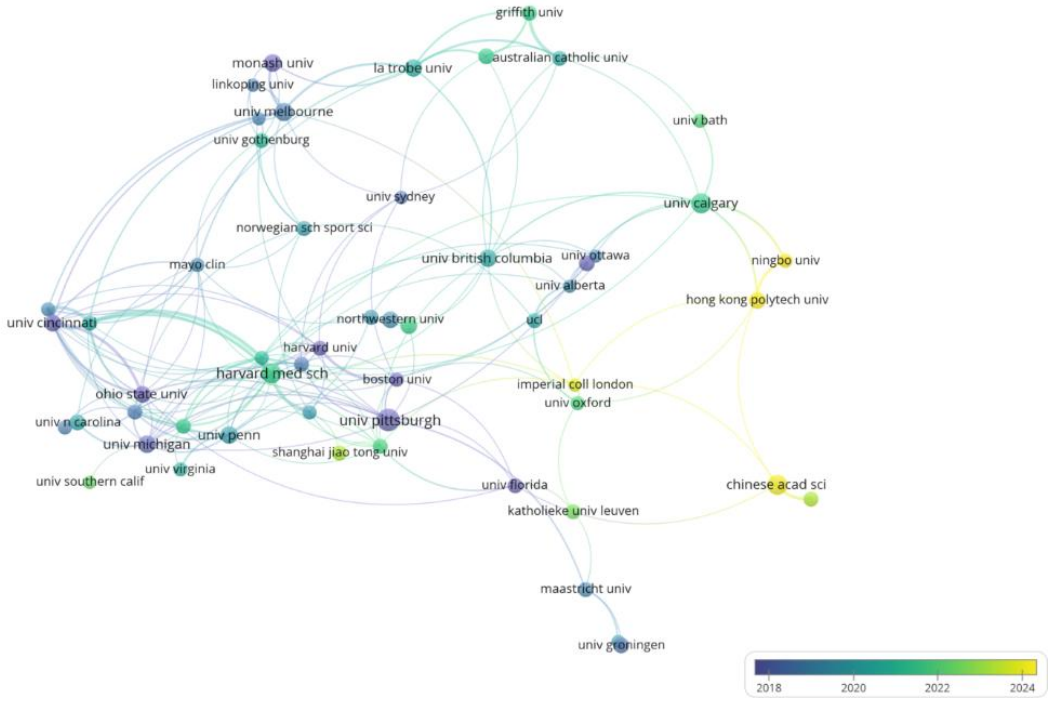


Figure 7) Collaboration among organisations

Author Contributions

At the author level, the analysis indicates that a core group of researchers contributes significantly to the development of this field. As illustrated in Figures 8, the leading authors demonstrate both high productivity and strong citation impact, with Myer Gregory D emerging as the most prolific contributor, with 18 publications. Co-authorship networks (Figure 9) reveal dense collaboration clusters, suggesting the presence of well-established research communities within the domain. In addition, the overlay analysis (Figure 10) indicates that newer authors, such as Ayramo Sami, are increasingly entering the field, contributing to its ongoing expansion and diversification.

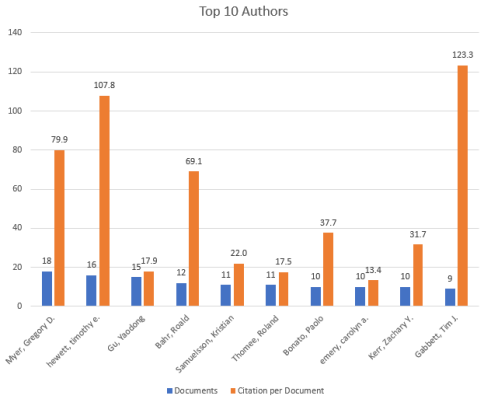


Figure 8) Top authors (Publications and citations)

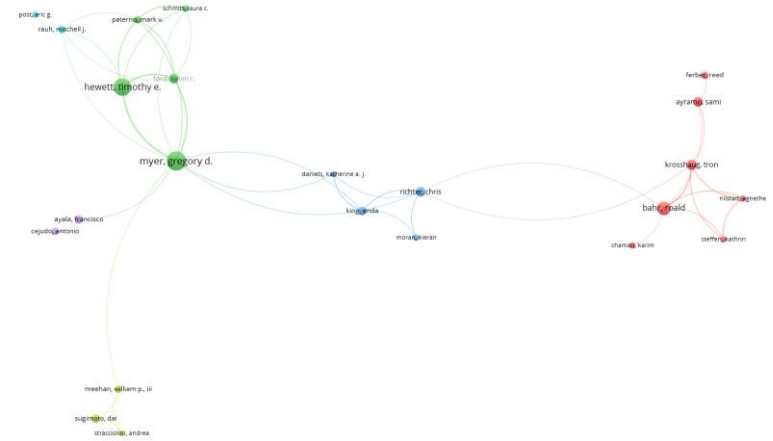


Figure 9) Authors' network

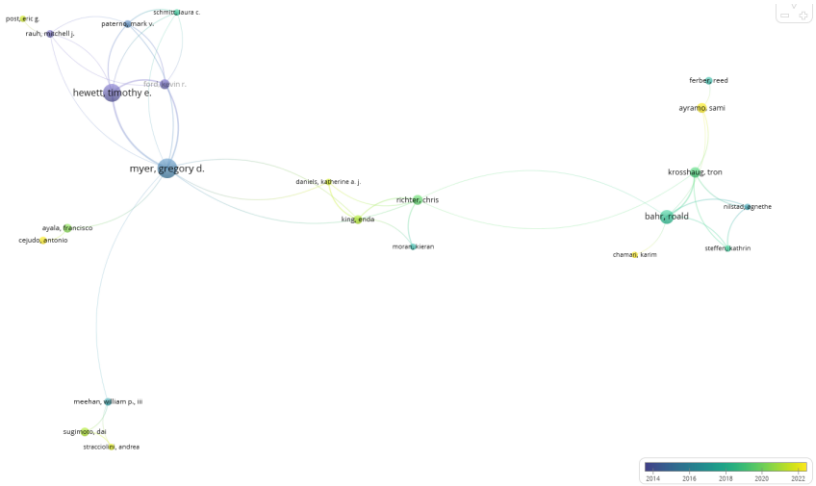


Figure 10) Authorship overlay analysis

3.3 Thematic Structure

The table below presents the top 20 keywords extracted from the dataset. The analysis of keyword frequencies and co-occurrence patterns provides insight into the conceptual structure of the field. The most frequently occurring keywords indicate that research is primarily centred on injury prediction and risk assessment, rehabilitation and recovery, machine learning models, and the use of wearable sensors and biomechanics. Together, these themes highlight the interdisciplinary nature of the field, reflecting the integration of data-driven approaches with sports science and health-related applications.

Table 1) Top 20 keywords

Keyword	Occurrences	Keyword	Occurrences
Machine Learning	495	Injury Prevention	191
Rehabilitation	446	Deep Learning	188
Biomechanics	328	Anterior Cruciate Ligament	174
Injury	318	Epidemiology	172
Wearable Sensors	283	Risk-factors	169
Sports	272	Prevention	158
Risk	259	Classification	157
Performance	245	Exercise	144
AI	234	Gait	144
Reliability	207	Kinematics	139

The keyword co-occurrence network (Figure 11) shows that "machine learning" functions as a central node connecting multiple research areas, highlighting its role as the core methodological foundation of the field. For clarity and improved readability of the visualisation, larger nodes, particularly the "machine learning" node have been intentionally scaled down in the figure.

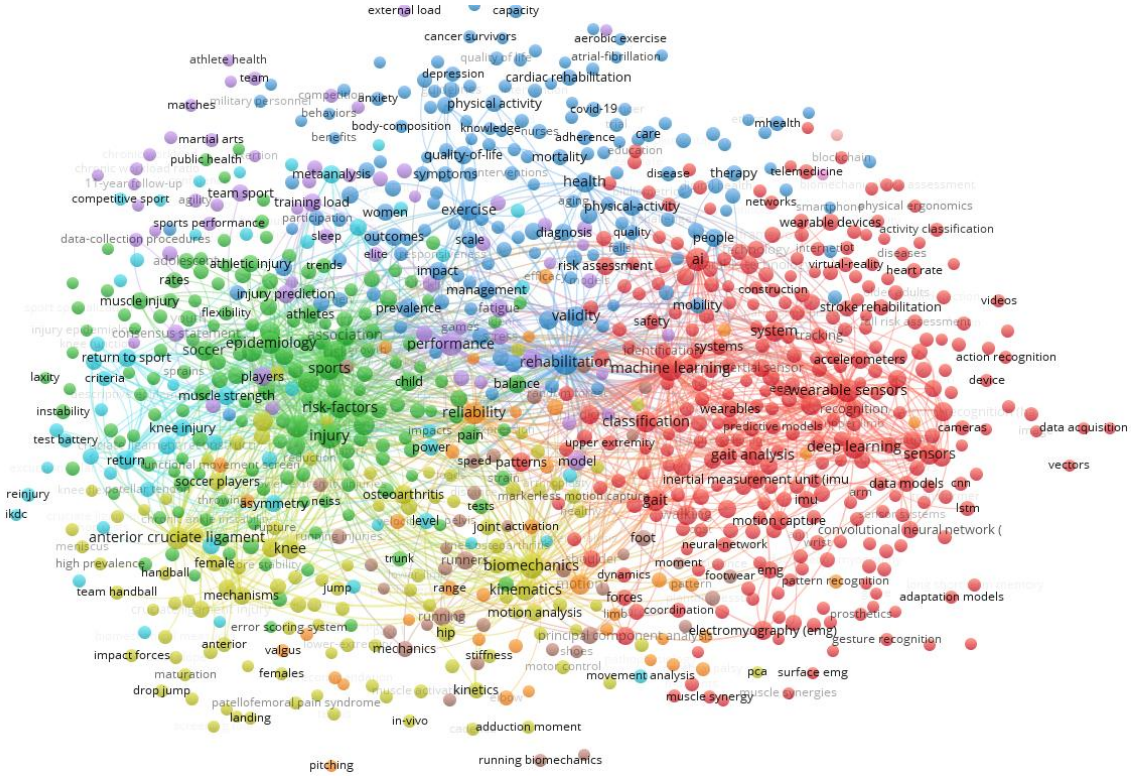


Figure 11) Keyword co-occurrence network

Additional visualisations (Figure 12) indicate that machine learning is strongly connected with interdisciplinary concepts such as biomechanics, performance monitoring, and physiological data, highlighting the close integration of data science and sports science within this research domain.

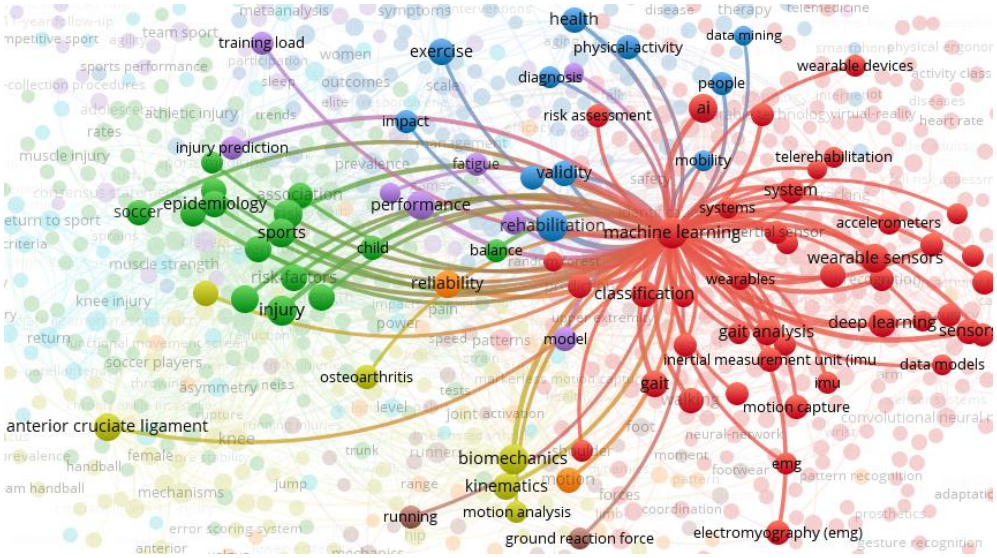


Figure 12) Studies around machine learning associated with other keywords

Temporal keyword analysis around ML (Figure 13), further indicates a clear shift in research focus over time. Earlier studies were more centred on sport-specific applications, often addressing particular disciplines or contexts like physical activity, running, and exercise, whereas more recent research places greater emphasis on generalisable predictive models and data-driven approaches, such as AI, data models, and injury prediction. This transition reflects the increasing maturity of the field, with a growing focus on developing scalable methods that can be applied across multiple sports and broader performance and injury contexts.

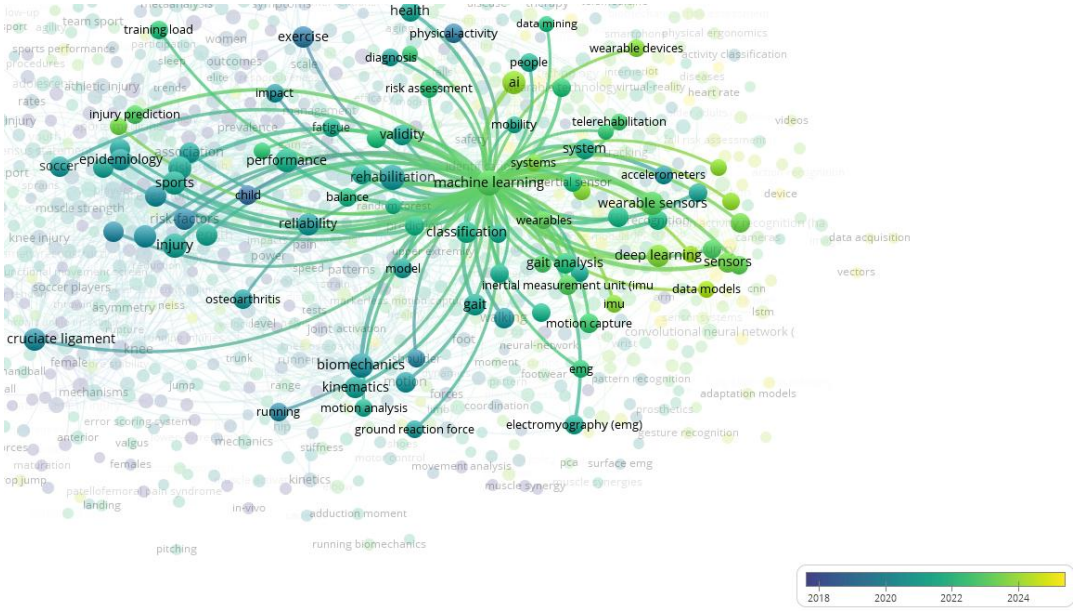


Figure 13) Overly keyword occurrence between machine learning and other keywords during time

3.4 Emerging Research Fronts

The overlay visualisation of keywords (Figure 14) reveals several emerging research areas that have gained prominence in recent years, including predictive models, wearable sensor technologies, explainable AI, real-time monitoring systems, and data-driven rehabilitation. These topics, such as explainable AI are associated with more recent time periods in the overlay map, indicating their position at the forefront of current research activity.

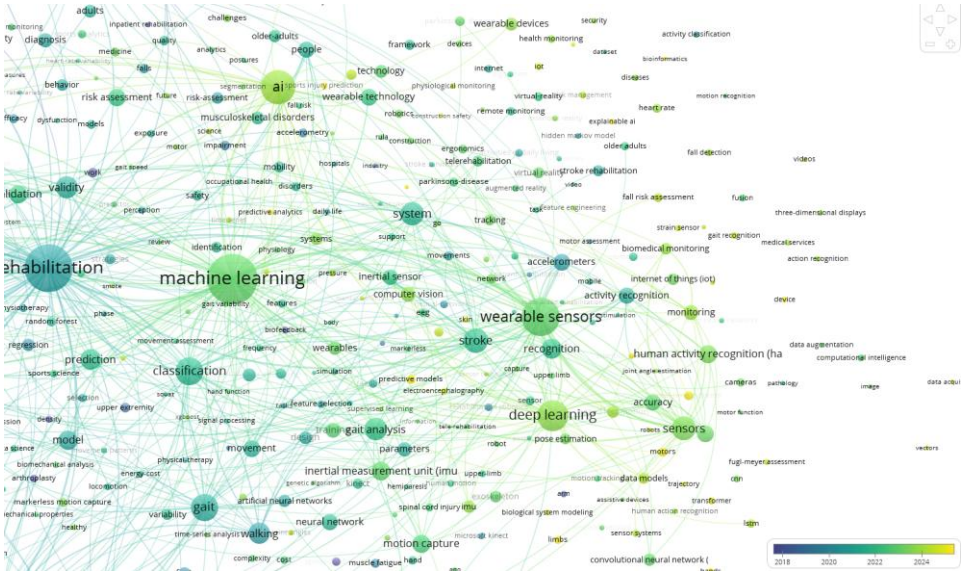


Figure 14) Overlay visualisation of keyword trends

4. DISCUSSION

This study provides a structured understanding of AI-driven sports injury research and shows that the field is moving from an emerging topic toward a more established research area. The rapid growth after 2015 is consistent with the increasing use of wearable technology, athlete monitoring systems, and predictive analytics in sport. For example, recent work has shown that wearable sensor data can support workload monitoring, return-to-play decisions, and predictive modelling for athlete health and safety, although current injury prediction models still face limitations due to small samples, low injury rates, and the complexity of injury mechanisms (Seshadri et al., 2021; Van Eetvelde et al., 2021). The country and institutional results show that research activity is concentrated among a limited number of leading countries and organisations, while collaboration networks indicate growing international connectivity. This pattern is important because AI-driven injury prediction requires expertise from several domains, including sports science, biomechanics, medicine, and data science. Therefore, the collaboration clusters observed in this study suggest that the field is developing through cross-disciplinary and cross-national knowledge exchange rather than isolated research activity.

The thematic findings also show a clear transition from sport-specific applications toward broader machine learning and predictive modelling approaches. This suggests that the field is maturing from isolated case-based studies toward more generalisable models that can be tested across different sports, populations, and injury contexts. This is consistent with recent evidence showing that machine learning has potential for sport injury prediction and prevention, but that stronger validation, clearer reporting, and more robust datasets are still needed before these models can be confidently adopted in practice (Van Eetvelde et al., 2021). The emergence of wearable technologies, real-time monitoring, explainable AI, and data-driven rehabilitation further indicates a practical shift in the field. For athlete health decisions, prediction alone is not sufficient; coaches, clinicians, and sport organisations also need models that are understandable, transparent, and actionable. Recent explainable AI literature similarly emphasises that transparency and interpretability are essential in high-stakes health-related decision-making, where errors may have serious consequences and trust in AI systems is critical (Sadeghi et al., 2024). Methodologically, this study also reinforces the value of bibliometric analysis for mapping a rapidly expanding and interdisciplinary field. Bibliometric analysis is useful for handling large volumes of scientific data, identifying collaboration patterns, detecting emerging topics, and mapping the intellectual structure of a domain (Donthu et al., 2021). However, as argued in prior research impact work, bibliometric evidence should be complemented with other methods when assessing practical, social, or policy impact, because publication and citation patterns do not fully capture real-world adoption or stakeholder value (Kamalpour et al., 2025).

5. CONCLUSIONS

This study provides a bibliometric overview of AI-driven sports injury prediction research, highlighting rapid growth since 2015, increasing global collaboration, and a shift from sport-specific studies toward generalisable, data-driven models. Emerging trends such as wearable technologies, explainable AI, and real-time monitoring point to a growing emphasis on practical, decision-support applications. Building on these findings, future

research should prioritise the development of interpretable AI models, the integration of real-time and multi-modal data, stronger cross-disciplinary collaboration, and the validation of models in real-world settings to ensure practical relevance and impact. From a policy perspective, the study provides evidence to support strategic research funding, international collaboration, and data-driven injury prevention initiatives. Bibliometric analysis, enables policymakers to identify research priorities, monitor knowledge development, and align investments with high-impact areas. Overall, this study offers a structured understanding of the field and a foundation for advancing both research and practice in AI-driven sports injury prevention.

References

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied computing and informatics, 15*(1), 27-33.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research, 133*, 285-296.

Kalkhoven, J. T., Watsford, M. L., Coutts, A. J., Edwards, W. B., & Impellizzeri, F. M. (2021). Training load and injury: causal pathways and future directions. *Sports Medicine, 51*(6), 1137-1150.

Kamalpour, M., Nili, A., & Yigitcanlar, T. (2025). Assessing Research Impact: A Framework and Implications for Management Information Systems Research.

Maffulli, N., Longo, U. G., Gougoulias, N., Caine, D., & Denaro, V. (2011). Sport injuries: a review of outcomes. *British medical bulletin, 97*(1), 47-80.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics, 106*(1), 213-228.

Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., & Hussain, S. (2024). A review of explainable artificial intelligence in healthcare. *Computers and Electrical Engineering, 118*, 109370.

Seshadri, D. R., Thom, M. L., Harlow, E. R., Gabbett, T. J., Geletka, B. J., Hsu, J. J., Drummond, C. K., Phelan, D. M., & Voos, J. E. (2021). Wearable technology and analytics as a complementary toolkit to optimize workload and to reduce injury burden. *Frontiers in sports and active living, 2*, 630576.

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538.

Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of experimental orthopaedics, 8*(1), 27.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational research methods, 18*(3), 429-472.

MATH Sport 2026

18th Australasian Conference on Mathematics and Computers in Sport | 1-3 July 2026

DAY THREE - FRIDAY 3 JULY 2026

Morning at Sofitel | Afternoon visit to Bond University Sports Precinct

25-minute talks: 20 minutes + 5 minutes Q&A. 15-minute talks: 12 minutes + 3 minutes Q&A. ★Powerplay contribution, @ Could not attend

The first page of each abstract or paper is shown in parentheses after each session title

09:00-11:10

TIME	MIN	AREA / PRESENTER	SESSION TITLE – Steven Azzopardi
09:00	30	AGM and Committee	Welcome
09:30	25	Running William Sutton	Pace profiling in middle-distance running: using functional data and clustering to understand how races unfold (199)
09:55	25	Athletes Tim Heazlewood	The use of mathematical models to predict athletic performance in men's and women's 100m, 200m, 400m, long jump, high jump, shot put and discus at the 2026 Glasgow Commonwealth Games (200)
10:20	25	Athletes Madison Stibbe	Bayesian network inference with uncertainty: modelling athlete progression (214)
10:45	25	MORNING TEA – POWERPLAY – Chair Marc West	
11:10	25	Concussion Hanem Ellethy	A decade of deep learning in sports-related concussion diagnosis: a review of neuroimaging approaches (2016-2026) (215)

11:35 ONWARDS

TIME	MIN	AREA / PRESENTER	SESSION TITLE
11:35	15 ★	Mental Health Rania Shibl & Shazi Saremi	Quantitative modeling of mental health improvement through aerial sports participation and nutrition Awareness (225)
11:50	15 ★	Golf Ant Bedford	Measuring excitement in the closing stages of the Masters golf (234)
12:05	15 ★	Powerplay TBC	Quiz + Tipping
12:20	15 ★	Yoga Tristan Barnett	Could Roger Federer have won more Grand Slams by following a yoga process? (242)
	@	Swimming Ray Stefani	Understanding how today's Olympic swimming champions maintain the velocity of past champions while swimming twice as far supporting future improvements
12:35	40	WORKING LUNCH AGM - AGM and Committee – Tim H	
13:15	45	Travel to Bond	Travel to Bond University - visit Sports Precinct. Car pool; 2:00 pm arrival.
14:00	120	Bond University	Visit the Bond University Sports Precinct
16:00		Leave Bond - Conference ends	

UniSC | BOND UNIVERSITY

PACE PROFILING IN MIDDLE-DISTANCE RUNNING: USING FUNCTIONAL DATA AND CLUSTERING TO UNDERSTAND HOW RACES UNFOLD

William Sutton [a,b,d] , Kerrie Mengersen [a,b] , Mitchell Mooney [c] , Paul Wu [a,b]

[a] *Centre for Data Science, Queensland University of Technology – Brisbane, Australia*
[b] *School of Mathematical Sciences, Queensland University of Technology – Brisbane, Australia*
[c] *Australian Athletics*
[d] *Corresponding author: wj.sutton@hdr.qut.edu.au*

Abstract
A key issue observed in major athletics championship middle-distance races is that athletes often don't achieve their season best performances, with the greater reliance on strategy and tactics required in championship racing seen as a possible contributing factor. To date, pace profiles which describe how an athlete exerts their effort across a race have been described by the shape of the speed-distance graph, often using terms like "all-out", "positive", "negative", "u-shaped" or "j-shaped" to classify the pace profile (Casado et al., 2021). Whilst this explains generally how races can unfold, it fails to capture some of the nuances within a race, for example different distances, genders or race types (paced vs unpaced). This research analyses middle-distance races across both major championship races (Olympic Games and World Championships) and regular season races (Diamond League) to better understand the pacing profiles that appear in these types of races and which pace profile types might lead to better athlete outcomes. The race split time series data was modelled as functional data, using functional Principal Component Analysis to map it to lower dimensional Principal Components (PCs) with associated mean profiles of pacing. Individual athlete performances were clustered based on their PC scores using Gaussian Mixture Modelling to identify different pace profiles which existed across the dataset. We found four major clusters of pacing profiles, and analysed contexts associated with them for the men's and women's 800m event. For instance, major championship races often differ from regular season races with different profiles seen for the two types of races. This information can be used by athletes and coaches to support training and strategic decisions, providing a better understanding of which pace profiles have historically led to better outcomes.

Keywords: Pace profile, middle distance running, functional data, clustering

Acknowledgements
We wish to thank Australian Athletics as the industry partner for this research. This project is made possible by CSIRO's Next Generation Artificial Intelligence Graduates Program (GA221787) funded by the Australian Government.

References
Casado, A., Hanley, B., Jiménez-Reyes, P., & Renfree, A. (2021). Pacing profiles and tactical behaviors of elite runners. *Journal of Sport and Health Science*, *10*(5), 537-549. https://doi.org/10.1016/j.jshs.2020.06.011

MATHEMATICAL MODELS TO PREDICT ATHLETIC PERFORMANCE IN MEN'S AND WOMEN'S 100M, 200M, 400M, LONG JUMP, HIGH JUMP, SHOT PUT AND DISCUS AT THE 2026 GLASGOW COMMONWEALTH GAMES

Ian T. Heazlewood

Sport Science Institute, Sydney, NSW, Australia
Corresponding author: edulearnsport2020@gmail.com

Abstract

Prediction of future athletic performances for men and women athletes in sprints, jumps and throes events have been based on mathematical models using regression linear and nonlinear curve estimation methods. It is a common practice in sport science and frequently based on plotting the changes in world or international records or averaging the performances of event finalists from previous Olympic Games and athletic World Championships. The regression curve estimation models predominantly indicated cubic nonlinear models were the best fitting model across sprint, jump and throw events for men and women. However, it is interesting to observe these methods have only been minimally applied to performance prediction ultising Commonwealth Games data. The research aims and design were to partially replicate the research of the Heazlewood (2024) Mathsport publication based on the Summer Olympic Games data. However, in this 2026 research data from 1938 to 2022 Commonwealth Games was used to predict averaged performances based on first six finalists in events of men's and women's 100m, 200m, 400m, long jump, high jump, shot put and discus at the 2026 Glasgow Commonwealth Games. The regression functions assessed to best fit the data were linear, logarithmic, inverse, quadratic, cubic, compound, power, sigmoidal, growth, exponential and logistic. The results indicated men 100m was cubic R^2 =.926, women 100m cubic R^2=.920, men 200m cubic R^2=.737, women 200m cubic R^2=.679, men 400m cubic R^2=.708, women 400m cubic R^2=.736, men long jump cubic R^2=.818, women long jump cubic R^2=.877, men high jump cubic R^2=.720, women high jump cubic R^2=.904, men shot put cubic R^2=.916, women shot put cubic R^2=.866, men discus cubic R^2=.941 and women discus cubic R^2=.775. For all events $p<.001$ and small residuals. The mathematical models derived, that is the cubic regression functions are very similar and slightly less accurate to the models derived for the 2024 Paris Olympic Games. Predicted performances based on the cubic functions derived for 2026 Commonwealth Games are 100m men 10.11s, 100m women 11.27s, 200m men 20.31s, 200m women 22.26s, 400m men 44.67s, 400m women 50.64s, men long jump 8.05m, women long jump 6.58m, men high jump 2.25m, women high jump 1.89m, men shot put 20.67m, women shot put 18.91m, men discus 62.32m and women discus 63.81m. Results predicted minimal change from the 2022 Commonwealth Games performances in these events.

Keywords: Predicting athletic event performance, Commonwealth Games, mathematical modelling, linear and non-linear regression curve estimation, men and women athletes.

1. INTRODUCTION

MATHEMATICAL MODELLING AND PREDICTING FUTURE ATHLETIC PERFORMANCE

The prediction of future athletic performance by athletes at the Olympic Games and World Athletic Championships is a recurring theme, as well as forming the basis for some stimulating discussions on the limits of human performance. Mathematical models in sport and exercise science should be based upon substantial evidence, rather they are based on the belief that records are made to be broken and that performances based on past experiences must continue to improve over time. Obviously this concept has limits and at some point the limits of human performance based on human biology and physiology will be achieved. The accessibility of data

in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From these analyses, changes in performance over time can be observed and predictions of future performance can be made utilising the process of mathematical extrapolation and interpolation.

A number of researchers have attempted to predict future performances by deriving and applying a number of mathematical statistical models based on past performances in athletics (Heazlewood and Lackey, 1996, 1998; Heazlewood, 2006a, 2006b, 2011a, 2011b, 2013a, 2013b; Heazlewood 2024; Heazlewood & Walsh, 2014a, 2014b, 2015). Prendergast (1990) applied the average speeds of world record times to determine a mathematical model for world records. The records or data used in the analysis spanned a 10 year period. Following his analysis, Prendergast (1990) raised the question of whether any further improvements can be expected or if the limits of human performance have been reached. The sports of athletics (Heazlewood and Lackey, 1996, 1998; Heazlewood, 2006a, 2006b, 2008, 2010, 2011a, 2011b, 2013a, 2013b, 2024 Heazlewood & Walsh, 2014a, 2014b, 2015; Walsh & Heazlewood, 2014) and swimming (Heazlewood & Lackey, 1996; Lackey and Heazlewood, 1998; Heazlewood, 2006a, 2006b) have been addressed in this manner.

Historically, the knowledge of future levels of sporting performance has been identified as beneficial in the areas of talent identification, both long and short term goal setting, and training program development based on the next level of expected future performance and the predicted limits to improving performance, and thus world records in the sport of athletics. In addition, expected levels of future performance are often used in the selection of national representative teams where performance criteria are explicitly stated in terms of athletics times and distances for example as required entry or qualification standards that are set at each Olympic Games and by World Athletics for entry at the Olympic Games and World Athletics Championships (International Olympic Committee (IOC), 2024).

The Commonwealth Games similarly set competition event standards to qualify to compete at games. A set of qualification standards and criteria are set by World Athletics prior to the Olympic Games (McAlister, 2022) and the standards for the Paris Olympic Games and the 2026 Commonwealth Games will be addressed in the methods section.

Péronnet and Thibault (1989) postulate that some performances, such as the men's 100m sprint is limited to the low 9 seconds, whereas, Seiler (referred to by Hopkins, 2000) envisages no limits on improvements based on data reflecting progression of records over the last 50 years. According to Seiler improvements per decade have been approximately 1% for sprinting, 1.5% for distance running, 2-3% for jumping, 5% for pole vault, 5% for swimming and 10% for skiing for male athletes, whereas female sprint times may have already peaked. The differences for males and females it is thought to reflect the impact of successful drugs in sport testing on females. Previous derived curve estimations that significantly fit the data have also displayed interesting findings as no one curve fits all the data sets.

Over the long term 1924 to 2020, Olympic event performances have improved incrementally sprint events in and significantly in jump and throwing events in the men's and women's 100m, 200m, 400m long jump, high jump, shot put and discus, which will be the Olympic events evaluated in this research (Olympic Results, International Olympic Committee (2024). The significant improvements in the jump and throwing events can be attributed to technical innovations as the hitch kick in long jump, Fosbury flop in high jump and rotational throwing technique in the shot put and discus, combined with increased strength and power training and increased mass and height of throwers adding to technique.

For men from 2004 to 2024 Olympics actual performances by the mean score of the top six finalist in each event (to represent overall athlete performance change) indicates improvements of 8.5 % from 10.82s to 9.90s men 100m (1924-2020), 9.5% from 21.92s to 19.83s men 200m (1924-2020), 8.5% from 48.35s to 44.24s men 400m (1924-2020), 16.22% from 7.09m to 8.24m men long jump (1924-2020), 24.87% from 1.89m to 2.36m men high jump (1924-2020), 34% from 14.59m to 22.12m men shot put (1924-2020), and 33.57% from 44.38m to 66.81m men discus (1924-2020).

For the women 11.95% from 12.30s to 10.83s women 100m (1928-2020), 12.96% from 25.22s to 21.95s women 200m (1948-2020), 8.65% from 54.09s to 49.41s women 400m (1964-2020), 25.32% from 5.53m to 6.93m women long jump (1948-2020), 25.16% from 1.59m to 1.99m women high jump (1928-2020), 33.40% from 13.06m to 19.61m women shot put (1948-2020), and 44.8% from 36.25m to 65.70m women discus (1928-2020).

Predictions of performance changes in athletics at the Commonwealth Games have not received the level of research as the Olympic Games, World Championships and changes in world records. The minimal research on the Commonwealth Games promoted hypotheses such as; are the trends in performances in the athletic events at Olympic Games and World Championships replicated at the Commonwealth Games for men and women

athletes in terms the mathematical models that best describe and predict these changes, are the trends similar for men and women in each event and will future performances increase, decrease or become static?

The current published research was very limited and no specific research could be located specific to regression modelling of performance changes over time (1930-2022) in the sprints, jumps and throws based on the Commonwealth Games. Observable research has examined changes over time considering best performance data, that is analysing gold medallist times and distances, however specific research was difficult to locate. I ran an AI search on Google (2026) which provided the following information using search descriptor "research papers trends in track and field performances at the commonwealth games". References cited by Google AI not specific to Commonwealth Games at;
(https://www.google.com/search?q=research+papers+trends+in+track+and+field+performances+at+the+commonwealth+games&sca_esv=f5c102df0dd72ac3&rlz=1C1CHBD_en-GBAU904AU912&biw=1024&bih=475&ei=bffeaeW9Hs-qseMPzZ6h8Qg&ved=0ahUKEwilqMqX5-6TAxVPVWwGHU1PKI44KBDh1QMIEQ&oq=research+papers+trends+in+track+and+field+performances+at+the+commonwealth+games&gs_lp=Egxnd3Mtd2l6LXNlcnAiUHJlc2VhcmNoIHBhcGVycyB0cmVuZHMgaW4gdHJhY2sgYW5kIGZpZWxkIHBlcmZvcm1hbmNlcyBhdCB0aGUgY29tbW9ud2VhbHRoIGdhbWVzSABQAFgAcAB4AJABAJgBAKABAKoBALgBDMgBAPgBAZgCAKACAJgDAJIHAKAHAHALIHALgHAMIHAMgHAIAIAA&sclient=gws-wiz-serp.

The following information is a direct quotation (Google AI Overview, 2026) and is based on inference from other similar information derived from Olympic Games and Athletic World Championships.
"Research into track and field performance trends at the Commonwealth Games and similar elite multi-sport events reveals a general trend of performance improvements up to the 1980s or 1990s, followed by a plateau or subtle decline in recent decades, with significant, lingering effects from technological advancements in equipment.
MAJOR TRENDS IN TRACK AND FIELD PERFORMANCE
Performance Plateau: Research suggests that track and field performances have largely reached an asymptotic level, with many peak performances occurring 1980–1999.
Recent Declines: Comparing the 1980–1999 period to 2000–2019, many events saw a decline in the best performance (2.56% for men, 1.67% for women), suggesting a peak rather than a plateau."

In summary, nonlinear regression analyses applied to extensive data from Olympic Games, World Championships and World athletic Top List athletes year by year data for all athletic events provides accurate models that describe and predict performances in the sprint, jump and throw events. The research question is do these trends also what is occurring in these events at the Commonwealth Games? The Google AI analysis indicates this trend should be replicated, however it forgot to mention that the majority of the predictive mathematical models were based on cubic regression solutions.

TRENDS IN CURVE ESTIMATION AND MATHEMATICAL FUNCTIONS

The curve estimation regression functions derived by Heazlewood and Lackey (1996) to predict performance in the men's and women's produced mix results across the sprint and jump events. In athletics for the men's events the mathematical functions (Heazlewood and Lackey, 1996) that displayed the best goodness of fit were 100m inverse, 400m sigmoidal, long jump cubic and the high jump displayed four functions (compound, logistic, exponential and growth). In the women's events the regression functions that displayed the best goodness of fit were 100m cubic, 400m sigmoidal, long jump inverse and the high jump displayed four functions (compound, logistic, exponential and growth).

The curves that fit the data have displayed interesting findings as no one curve fits all the data sets and men and women displayed some different functions linked to gender as men's 100m was inverse whereas women's was cubic, men's long jump cubic and women's long jump inverse. The function for the 400m and high jump were the same for men and women. This may indicate that different events and gender responses are dependent upon different factors that are being trained differently or factors underpinning performance evolving in differences in men and women exercise physiology and biomechanics. This has resulted in different curves or mathematical functions that reflect these improvements in training or phylogenetic changes over time. However, at some point in the future when time catches-up with the actual performance, then how accurately the predictive models reflect reality can be assessed.

The Heazlewood (2006a, 2006b) study to a degree replicated the findings in the 1996 study (Heazlewood & Lackey, 1996), where the functions and trends for men and women in the 100m, 400m, long jump and high jump were very similar. The study was evaluating the predictions of the 2000 Sydney and 2004 Athens Olympic Games. It is interesting to observe the models derived in the 2006a and 2006b research moderately overestimated performance for men's and women's 100m although actual times did improve from 2000 to 2004, overestimated performance in the men's and women's 400m, moderately overestimated performance in the men's long jump

and significantly in the women's and moderately overestimated performance in the men's and women's high jump. Table 1 indicates the athletic events, type of mathematical function, equations and R^2 values from the Heazlewood (2024) research for 2024 Paris Olympics. Where the regression model for the men is a cubic regression function across 100m, 200m, 400, high jump, long jump, shot put and discus all displaying high explained variance (R-square values) and accurate curve fitting.

For women there were a mixture of regression solutions where the best fit for the 100m, high jump and discus was cubic with high R^2 values, sigmoidal for 200m, 400m and shot put and inverse for long jump. This results suggest different progression functions and underlying factors for different event in women athletes, however not in male athletes as all function/models were cubic suggesting a common progression process.

However, the ability to predict performances at the 2000, 2004, 2008, 2012 and 2024 Olympic Games for the men's 100m, 400m, long jump and high jump, based on the 1924 to 2012 data, were moderately accurate with low to moderate residual error.

The prediction performance in the throwing events of shot put and discus for men and women (Heazlewood & Walsh, 2014a, 2014b) produced results different from the earlier sprint and jump findings. Note the 2014a and 2014b studies evaluated all throwing events, the shot, discus, javelin and hammer, however the focus will be the findings for the shot put and discus to be consistent with this current research. The International Olympic Committee (IOC) athletic results was the source for the data set (IOC, 2014) when this study was completed and produced descriptive data and descriptive graphical analysis to indicate trends in World records for the men's throwing events, however the analyses were not based on mathematical predictive modelling. The data set in this research for men was 1936 to 2012 for shot put and discus, and for women was 1948-2012 for shot put and 1928-2012 for discus. The curve estimation regression functions for the men both for shot put and discus were cubic functions with high explained variance (shot put R^2= .965 and discus R^2= .972) and low residual error with inflexion points both suggesting performance decreases in 2016 and 2020 and subsequently to confirmed for the discus in 2016 and 2020, however not for the shot put as the 2016 prediction was accurate, however the 2020 result displayed a significant improvement in performance and not predicted in the cubic model.

The women's shot put and discus conformed to a cubic function as the best fit to the data and similarly displayed high explained variance (shot put R^2= .918 and discus R^2= .925) and low residual error. Once again, inflexion points were suggesting performance declines in these two events in the near future and this is what occurred in both events at the 2016 and 2020 Olympic. What is interesting based the results are shot put and discus events for both men and women competing at the Olympics displayed cubic functions that predicted performance plateaux and then performance declines and suggesting in these events the men and women shot put and discus throwers are approaching their performance limits.

The Heazlewood (2024) research predicting athletic performance at the 2024 Paris Olympics in 100m, 200m, 400m, high jump, long jump, shot put and discus events for men and women indicated increase in 100m performance in men and women, 200m a decrease in men and increase in women performance, 400m static for men and increase for women, high jump static for men and increase in women, long jump increase in both men and women performance, shot a decrease for men and increase for women and discus increase for men and decrease for women. It is important to highlight the Heazlewood (2024) data is the most recent and based on a larger set of Olympic Games data from 1928 to 2020.

The dominant and important research question is, can mathematical models based on nonlinear curve estimation, which have proven to be very successful in fitting and predicting past, current and future Olympic performances for event finals in athletics and swimming for men display equal effectiveness in predicting athletic performances in the men's throwing events of the shot put, discus, hammer and javelin at the 2026 Commonwealth Games, which are past, current and future performances?

2. RESEARCH PROBLEM, QUESTIONS AND HYPOTHESES

RESEARCH PROBLEM

The previous research that attempted to predict future athletic performance was based on linear regression and non-linear models in in the sport of athletics most curve estimation regression techniques found to be non-linear (Heazlewood and Lackey 1996; Heazlewood and Lackey, 1998; Heazlewood, 2006a, 2006b, 2024; Heazlewood & Walsh, 2011a; Heazlewood & Walsh, 2011b; Heazlewood & Walsh, 2014a; Heazlewood & Walsh, 2014b; Heazlewood & Walsh, 2015).

The research problem is the models developed in earlier research over 10 years ago still apply to current athletic events and enable prediction of past, current and future athletic events at the Commonwealth Games

such as Glasgow 2026, which will enable athletes to set realistic training and performance goals based on future performance standards to qualify for athletic events and aspire to achieving Commonwealth Games finals, which is where the medals are contested? The higher quality of predictive models will enable higher quality predictions of athletic performance by applying principles of sport science, mathematics and computing in sport is to understand and predict sports performance.

RESEARCH QUESTIONS

A number of research questions can be generated that will address the predominant theoretical issues. These are:

1. What is the best mathematical-statistical model that will best fit the data for each event in the Commonwealth Games?
2. Does one model fit all events?
3. Will the mean score based on the top six (finalists) better reflect actual changes in human performance that have occurred in men and women?
4. Will the models be identical for the different genders for each event?
5. Will the mathematical-statistical models enable the accurate prediction of future performances as well as past performances?

HYPOTHESES

Predicting Performances in Sprint, Jump and Throw Events at the 2026 Glasgow Commonwealth Games. Specifically in the 110m, 200m, 400m, high jump, long jump, shot put and discus.

1. Based on the most recent evidence (Heazlewood, 2024) the different events will have different models of best fit as they will be dependent on different human factors such as speed, power, endurance, speed endurance, strength and motor co-ordination that are represented by different mathematical-statistical functions which will display high explained variance across all events evaluated.
2. The different genders will have different models of best fit. Male athlete models will conform to the cubic regression model. Females will display a mix of models that represent the best fits for the different events and based on previous research should be cubic, sigmoidal and inverse models as a result of females only recently adopting more stressful physical training programs that will influence the response curves, the training and competition performance interaction over time and some events introduced from 1950.
3. The mathematical-statistical models will be both event and gender specific.
4. More of the variance will be explained and a better model fit will occur based on nonlinear mathematical functions.
5. The likelihood of absurd mathematical predictions will be reduced or non-existent as best fit models with high explained variance will be derived.

3. METHODS

SAMPLE AND EVENT SELECTION

The quality and quantity of athletes in the Commonwealth Games in the sport of athletics is not as high as in the Olympic Games and World Championships as evidenced by actual performances at these international championships. The differences in qualifying standards between the 2024 Olympic Games and 2026 Commonwealth Games highlight this quality of athlete concept. The following table 1 indicates the different qualifying event performance criteria between the two international competitions.

Event	Men 2024 Paris	Women 2024 Paris	Men Glasgow 2026	Women Glasgow 2026
100m	10.00s	11.07s	10.16-10.21s	11.17-11.34s
200m	20.16s	22.57s	20.54-20.61s	23.08-23.23s
400m	45.00s	50.95s	44.96-45.55	50.75-51.85
High Jump	2.33m	1.97m	2.21m	1.86-1.92m
Long Jump	8.27m	6.86m	7.88m	6.42m
Shot Put	21.50m	18.80m	19.42m	16.92m
Discus	67.20m	64.50m	59.56m	55.02m

Table 1: Entry Qualification Standard for The 2024 Paris Olympic Games and 2026 Glasgow Commonwealth Games Based On Actual Event Performances of Men and Women for 100m, 200m, 400m, High Jump, Long Jump, Shot Put And Discus.

The data sources and data sets were from the results of the Commonwealth Games (previously named the Empire Games prior to 1970) from 1938 to 2022 (Wikipedia, 2026). The games did not occur in 1942 and 1946 due to effects of World War II. Specially, athletes who finished 1 to 6 in the finals data were averaged to be more representative of overall performance of an event at that point in time.
The complete data set for men and women were sprint, jump and throw events of 100m, 200m, 400m, high jump, long jump, shot put and discus. All men's events in the 100m, 200m, 400m, high jump, long jump, shot put and discus were recorded from 1938 to 2022. Women's events were introduced at different times where 100m (1938-2022), 200m (1950-2022), 400m (1966-2022), high jump (1938-2022), long jump (1938-2022), shot put (1954-2022) and discus (1954-2022).

The sprint times were entered to the hundredth of seconds and distances for jumping and throwing events to the centimetre as provided by the data sources. From 1938 to 1966 to improve comparison between years in the 100m and 200m corrections were made for lack of automatic timing and distances as the 100 yards, 200 yards and 400 yards had to be converted to approximate 100m, 200m and 400m times. The conversion methods for yards to metres time.

- 100-To convert 100-yard time to a 100-meter time, the most common methods are to multiply the 100-yard time by 1.0936 for a constant speed conversion or add 0.9 seconds.
- 200-To convert a 220-yard time to a 200-meter time, the standard method used by track and field organizations is to subtract 0.1 seconds from the 220-yard time.
- 400m- To convert a 440-yard dash time to a 400-meter time, subtract 0.3 seconds from the 440y time.

Converting hand timing to automatic time.

- 100m and 200m – Add 0.24 seconds to hand time.
- 400m – Add 0.14 seconds.

Correction for assistive or resistive winds.

- 100m-A 100m sprint time is generally corrected for wind by adding approximately 0.10 to 0.12 seconds for every +1.0 m/s for assistive wind and subtracting 0.10 to 0.12 seconds for every -1.0 m/s for resistive wind (Mureika, 2026).
- 200m- A 200m sprint time is generally corrected for wind by adding approximately 0.10 to 0.14 seconds for every +2.0 m/s for assistive wind and subtracting 0.10 to 0.14 seconds for every -2.0 m/s for resistive wind (Mureika, 2026).

In events as 400m, high jump, shot put and discus no wind information is provided or considered in these events. In the long jump wind speed is provided, however no changes in scores occur in this event and is measured in case a specific record is broken. In this case the assistive wind speed must be 2m/s or less.

MODEL SELECTION AND DATA FIT CRITERIA

The data set was analysed utilising IBM SPSS Statistics-Version 26 (IBM, 2024). The following eleven curve estimation regression mathematical models were applied. An identical approach utilised in the evaluation of mathematical models that evaluate performances changes over time with Olympic Games (Heazlewood & Lackey, 1996; Lackey & Heazlewood, 1998; Heazlewood, 2006a, 2006b; Heazlewood, 2013a; Heazlewood & Walsh, 2014a; Walsh & Heazlewood; 2014b, Heazlewood, 2024), World Athletic Championships (Heazlewood; 2013b), top twenty performances for the major track and field events for years 1960 to 2014 for men and women senior outdoor athletes (Heazlewood & Walsh, 2015), and decline in Masters Athletic performance with age in track and field events (Heazlewood & Lackey, 1998). The evaluated regression mathematical models were linear, logarithmic, inverse, quadratic, cubic, power, compound, sigmoidal, logistic, growth and exponential.

GENERAL METHOD OF DETERMINING THE APPROPRIATE REGRESSIONS MODELS

To investigate the hypotheses of model fit and prediction, the eleven regression models were individually applied to each of the selected athletic and events. The regression equation that produced the best fit for each event, that is, produced the highest coefficient of determination (abbreviated as R^2), was then determined from these eleven equations. The specific criteria to select the regression equation of best were the magnitude of R^2, the significance of the analysis of variance alpha or p-value and the residuals.

1. The Coefficient of Determination

The coefficient of determination (R^2) is a measure of accuracy of the model used. A coefficient of determination of 1.00 indicates a perfectly fitting model where the predicted values match the actual values for each independent variable (Norušis, 1993; Garson, 2010; Hair et al., 2006). Where more than one model was able to be selected due to an equal R^2, the simplest model was used under the principle of parsimony, that is, the avoidance of waste and following the simplest explanatory model.

2. Residuals
The residuals are the difference between the actual value and the predicted value for each case, using the regression equation (Norušis, 1993; Garson, 2010; Hair et al., 2006) and the smaller the residual, the better the fit of the model. For each model the residuals were generated by the SPSS program. A large number of positive residuals indicate that the prediction is an over estimation (faster than the actual performance) and a large number of negative residuals indicates an underestimation (slower time than the actual performance).
3, Level of Significance
The level of significance, or p value, is a representation of the relationship between the model and the data. The smaller the p-value, the higher the level of significance and the greater the relationship where a small p-value indicates a small possibility that the closeness of the predicted values to the actual values due to chance is small.
4. Logical Acceptance Based on Extrapolations
The ability of the model to generate extrapolations that appear to be reasonable when compared to previous means was also taken into consideration. When a model generated extrapolations that appear to be inconsistent with the actual results this model was discarded and the model with the next highest coefficient of determination was selected. The models and mathematical function curve estimations finally selected were based on the preceding model evaluation criteria.
5. Applying the Model of Best Fit
After selection of the model to be used, according to the criteria previously stated, the equation of best fit was determined by applying the derived constants and coefficients to the generic formula for that model. Using this equation, a prediction of the mean result for the event at each Olympiad was calculated. At this stage, graphs representing the means of past and future performances for each event in each Olympiad were also generated in addition to predicted means using the appropriate regression equation. Final predictions have included prediction for 2026 Commonwealth Games, Glasgow.

4. RESULTS

ACTUAL AND PREDICTED PERFORMANCES AT 2026 COMMONWEALTH GAMES

The results for the curve estimation models for the different events are displayed in table 2. The men's model of best fit was for the cubic function for all events, which are the 100m, 200m, 400m, long jump, high jump, shot put and discus. This indicates a consistent mathematical trend across all these events in terms of performance change over time. The high R^2 values range from 0.708 for the men's high jump to 0.941 for the men's discus and all p-values are less than 0.001.

Event	Men Model	R-square	p-value	Women Model	R-square	p-value
100m	Cubic	0.926	<.001	Cubic	0.920	<.001
200m	Cubic	0.737	<.001	Cubic	0.679	<.001
400m	Cubic	0.708	<.001	Cubic	0.736	<.001
High jump	Cubic	0.720	<.001	Cubic	0.904	<.001
Long Jump	Cubic	0.818	<.001	Cubic	0.877	<.001
Shot Put	Cubic	0.916	<.001	Cubic	0.866	<.001
Discus	Cubic	0.941	<.001	Cubic	0.775	<.001

Table 2: Summary Table of Mathematical Models for Each Event and Gender for 2026 Commonwealth Games.

The women's model of best fit was for the cubic function for all events, which are the 100m, 200m, 400m, long jump, high jump, shot put and discus. This indicates a consistent mathematical trend across all these events in terms of performance change over time. The high R^2 values range from 0.679 for the women's 200m to 0.920 for the women's 100m and all p-values are less than 0.001. Evaluating each event by gender the following trends were observed.

For the 100m a cubic regression model was derived for both men and women and the predicted trend is performance stabilisation for the 2026 games as displayed in table 3. This indicates the 100m performance may have achieved its limit. 2026 performance predicted to be men 10.11s, whereas in 2022 it was 10.08s and 2026 performance for women 11.27s, whereas 2022 it was 11.12s and which represent marginal declines in performance. The root mean square error (RMSE) in the predictive model is low where 0.085s (0.8% error) for men and (1% error) for women.

Year	Men Corrected time(s) mean scores.	Predicted time(s) Cubic model Mean scores. R^2=0.926, p<.001	Residual(s)	Women Corrected time(s) mean scores.	Predicted time(s) Cubic model Mean scores. R^2=0.920, p<.001	Residual(s)
1938	11.04	11.19	-.15	12.54	12.57	-.03
1950	10.99	10.91	.08	12.27	12.19	.08
1954	10.84	10.83	.01	11.99	12.07	-.08
1958	10.90	10.75	.15	11.99	11.97	.02
1962	10.73	10.67	.06	11.92	11.87	.05
1966	10.57	10.60	-.03	11.92	11.78	.14
1970	10.51	10.54	-.03	11.66	11.70	-.04
1974	10.55	10.48	.07	11.51	11.62	-.11
1978	10.42	10.42	.00	11.53	11.55	-.02
1982	10.31	10.37	-.06	11.29	11.49	-.20
1986	10.38	10.32	.06	11.50	11.43	.07
1990	10.27	10.28	-.01	11.51	11.38	.13
1994	10.14	10.24	-.10	11.23	11.34	-.11
1998	10.01	10.21	-.20	11.26	11.31	-.05
2002	10.13	10.18	-.05	11.22	11.28	-.06
2006	10.21	10.16	.05	11.41	11.26	.15
2010	10.24	10.14	.10	11.47	11.25	.22
2014	10.12	10.12	.00	11.13	11.24	-.11
2018	10.21	10.11	.10	11.32	11.25	.07
2022	10.08	10.11	-.03	11.12	11.26	-.14
2026		10.11			11.27	

Table 3: Men and Women 100m Corrected Times Plus Best Fitted Model, Predicted Times And Residuals 1938 To 2022. Men RMSE=0.085s, Women RSME= 0.109s.

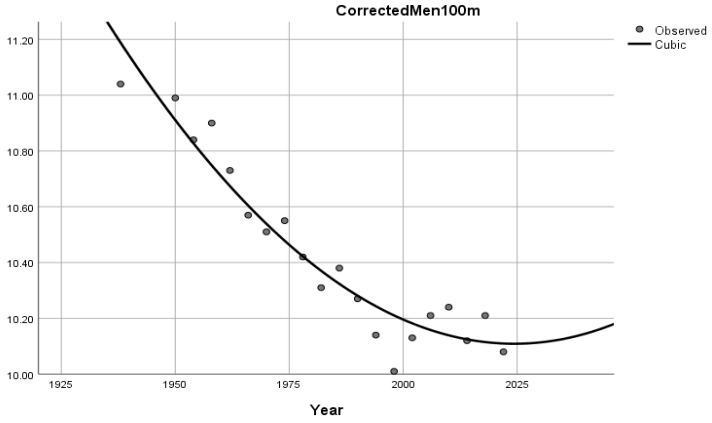


Figure 1. Graphically illustrates for the men's 100m the relationship between actual data points, the dots, and the line of best fit, which is the cubic regression function. X-axis 100m time in seconds and Y-axis years of competition.

For the 200m the cubic regression model was the best fit for both men and women as displayed in table 4. Again, a marginal increase in performance is predicted for men from 20.47s in 2022 to 20.31s in 2026, as well as a significant improvement is predicted for women from 22.84 in 2022 to 22.26 in 2026.The root mean square error in the predictive model is low where 0.085s for men and 0.109s for women. The RMSE for men is moderately low at 0.295s (1.4% error) and for women at 0.435s (1.9% error).

Year	Men Corrected	Predicted time(s)	Residual(s)	Women Corrected	Predicted time(s)	Residual(s)

	time(s) mean scores.	Cubic model Mean scores. R^2=0.737, p<.001		time(s) mean scores.	Cubic model Mean scores. R^2=0.679, p<.001	
1938	21.62	22.12	-.50		24.79	
1950	21.99	21.73	.26	24.85	24.45	.40
1954	21.99	21.61	.38	25.06	24.33	.73
1958	21.40	21.49	-.09	24.21	24.22	-.01
1962	21.84	21.38	.46	24.51	24.10	.41
1966	21.21	21.28	-.07	24.23	23.99	.24
1970	20.86	21.18	-.32	23.41	23.87	-.46
1974	21.02	21.08	-.06	23.06	23.76	-.70
1978	21.71	20.99	.72	23.17	23.64	-.47
1982	20.62	20.91	-.29	22.72	23.53	-.81
1986	20.85	20.83	.02	23.56	23.41	.15
1990	20.45	20.75	-.30	23.24	23.29	-.05
1994	20.49	20.68	-.19	22.69	23.18	-.49
1998	20.37	20.62	-.25	22.96	23.06	-.10
2002	20.30	20.56	-.26	22.71	22.95	-.24
2006	20.58	20.51	.07	22.99	22.83	.16
2010	20.65	20.46	.19	23.44	22.72	.72
2014	20.34	20.41	-.07	22.63	22.60	.03
2018	20.53	20.37	.16	22.51	22.49	.02
2022	20.47	20.34	.13	22.84	22.37	.47
2026		20.31			22.26	

Table 4: Men and Women 200m Corrected Times Plus Best Fitted Model, Predicted Times And Residuals 1938 To 2022. Men RMSE=0.295s, Women RSME= 0.435s.

For the 400m the cubic regression model was the best fit for both men and women as displayed in table 5. A slight improvement in the 400m is predicted where is was 45.19s in 2022 and predicted to be 44.67s in 2026. For women an improvement is also predicted to occur from 51.04 in 2022 to 2026. The Men RMSE values are low for the 400m event as men were 0.459s (1.4% error) and women were 0.476s which represents (2% error).

Year	Men Mean score (s)	Predicted time(s) Cubic model Mean scores R^2=0.7080, p<.001	Residual(s)	Women Mean score (s)	Predicted time (s) Cubic model Mean scores R^2=0.735, p<.001	Residuals(s)
1938		47.71			54.68	
1950		47.29			54.13	
1954	47.82	47.15	.67		53.95	
1958	47.15	47.02	.13		53.76	
1962	47.09	46.88	.21		53.58	
1966	46.17	46.74	-.57	54.07	53.40	.67
1970	45.73	46.60	-.87	53.68	53.21	.47
1974	46.32	46.46	-.14	52.73	53.03	-.30
1978	46.80	46.33	.47	52.97	52.85	.12
1982	46.78	46.19	.59	52.57	52.66	-.09
1986	46.59	46.05	.54	51.86	52.48	-.62
1990	45.24	45.91	-.67	52.38	52.29	.09
1994	45.49	45.77	-.28	51.11	52.11	-1.00
1998	45.00	45.63	-.63	51.30	51.93	-.63
2002	45.25	45.50	-.25	52.12	51.74	.38
2006	45.24	45.36	-.12	51.31	51.56	-.25

2010	45.60	45.22	.38	51.81	51.38	.43
2014	45.05	45.08	-.03	51.80	51.19	.61
2018	45.13	44.94	.19	50.91	51.01	-.10
2022	45.19	44.81	.38	51.04	50.82	.22
2026		44.67			50.64	

Table 5: Men and Women 400m Times Plus Best Fitted Model, Predicted Times And Residuals 1938 To 2022. Men RMSE=0.459s, Women RSME= 0.476s.

The high jump for both men and women displayed a cubic regression as the model of best fit as displayed in table 6. The prediction for men was a marginal increase in performance from 2.23m in 2022 to 2025 in 2026. The prediction for women indicated a marginal decline in performance from 1.92m to 1.89m. The RMSE for men was 0.035m (error 1.57%) and for women 0.038m (error 1.98%).

Year	Men Mean score (m)	Predicted height(m) Cubic model Mean scores R^2=0.720, p<.001	Residual(m)	Women Mean score(m)	Predicted height(m) Cubic model Mean scores R^2=0.904, p<.001	Residual (m)
1938	1.88	1.83	.05	1.55	1.48	.07
1950	1.95	1.95	.00	1.54	1.60	-.06
1954	1.97	1.99	-.02	1.59	1.63	-.04
1958	2.01	2.02	-.01	1.66	1.67	-.01
1962	2.05	2.05	.00	1.73	1.70	.03
1966	2.00	2.08	-.08	1.69	1.73	-.04
1970	2.09	2.11	-.02	1.69	1.75	-.06
1974	2.12	2.14	-.02	1.79	1.78	.01
1978	2.16	2.16	.00	1.84	1.80	.04
1982	2.22	2.18	.04	1.84	1.82	.02
1986	2.18	2.20	-.02	1.89	1.84	.05
1990	2.27	2.21	.06	1.88	1.85	.03
1994	2.30	2.23	.07	1.91	1.87	.04
1998	2.27	2.24	.03	1.89	1.88	.01
2002	2.23	2.25	-.02	1.87	1.89	-.02
2006	2.23	2.25	-.02	1.85	1.89	-.04
2010	2.26	2.26	.00	1.85	1.90	-.05
2014	2.24	2.26	-.02	1.91	1.90	.01
2018	2.28	2.26	.02	1.90	1.90	.00
2022	2.23	2.26	-.03	1.92	1.90	.02
2026		2.25			1.89	

Table 6: Men and Women high jump distances plus best fitted model, predicted times and residuals1938 to 2022.Men RMSE=0.035m: women RSME= 0.038m.

The long jump again displayed the cubic regression model as the best fitting model and displayed in table 7. The men were predicted to display a marginal increase in performance from 8.03m in 2022 to 8.05m in 2026, whereas the women were predicted to display s significant decrease in performance from 6.82m in 2022 to 6.58m in 2026. The RSME values were men 0.133m (error 1.7%) and women 0.131m (error 1.9%). The significant decline in women suggests a negative turning point has been attained and possibly performance limits achieved.

Year	Men Mean score (m)	Predicted distance(m) Cubic model Mean scores R^2=0.818, p<.001	Residual(m)	Women Mean score (m)	Predicted distance(m) Cubic model Mean scores R^2=0.877, p<.001	Residuals(m)
1938	7.17	7.05	.12	5.47	5.37	.10
1950	7.09	7.32	-.23	5.59	5.73	-.14

1954	7.25	7.40	-.15	5.67	5.84	-.17
1958	7.35	7.47	-.12	5.87	5.94	-.07
1962	7.73	7.54	.19	6.02	6.03	-.01
1966	7.78	7.61	.17	6.16	6.12	.04
1970	7.81	7.67	.14	6.34	6.19	.15
1974	7.75	7.73	.02	6.35	6.27	.08
1978	7.93	7.78	.15	6.42	6.33	.09
1982	7.70	7.83	-.13	6.66	6.39	.27
1986	7.62	7.87	-.25	6.32	6.44	-.12
1990	8.10	7.91	.19	6.56	6.49	.07
1994	7.94	7.95	-.01	6.41	6.53	-.12
1998	7.94	7.98	-.04	6.51	6.56	-.05
2002	7.89	8.00	-.11	6.49	6.58	-.09
2006	8.01	8.02	-.01	6.63	6.60	.03
2010	8.01	8.03	-.02	6.40	6.61	-.21
2014	8.06	8.05	.01	6.46	6.61	-.15
2018	8.15	8.05	.10	6.66	6.61	.05
2022	8.03	8.05	-.02	6.82	6.59	.23
2026		8.05			6.58	

Table 7: Men and Women Long Jump Distances Plus Best Fitted Model, Predicted Times And Residuals 1938 To 2022.Men RMSE=0.133m, Women RSME= 0.131m.

The shot put again conformed to a cubic regression model as the best fit with both men and women having high explained performance variance (men R^2=0.916, women R^2=0.866) and displayed in table 8. In this event the men were predicted to display performance decline 20.83m in 2022 to 20.67m in 2026, whereas the women were predicted to display significant improvement from 18.16m in 2022 to 18.91m in 2026. RMSE values indicated for men 0.597m (error 2.85%) for women 0.584m (error 3.22%).

Year	Men Mean score(m)	Predicted distance(m) Cubic mean scores R^2=0.916, p<.001	Residual(m)	Women Mean score(m)	Predicted distance(m) Cubic mean scores R^2=0.866, p<.001	Residuals(m)
1938	13.64	13.47	.17		12.64	
1950	13.60	14.97	-1.37		13.49	
1954	15.02	15.44	-.42	12.72	13.78	-1.06
1958	16.08	15.88	.20	13.73	14.06	-.33
1962	17.08	16.31	.77	13.79	14.35	-.56
1966	17.36	16.72	.64	15.15	14.63	.52
1970	17.78	17.11	.67	14.90	14.92	-.02
1974	18.60	17.49	1.11	15.01	15.20	-.19
1978	18.20	17.84	.36	16.13	15.49	.64
1982	17.52	18.18	-.66	16.75	15.77	.98
1986	18.07	18.50	-.43	17.48	16.06	1.42
1990	17.99	18.80	-.81	16.32	16.34	-.02
1994	18.86	19.08	-.22	16.89	16.63	.26
1998	19.32	19.35	-.03	16.77	16.91	-.14
2002	19.52	19.59	-.07	16.68	17.20	-.52
2006	19.16	19.82	-.66	17.74	17.48	.26
2010	19.84	20.02	-.18	17.46	17.77	-.31
2014	20.31	20.21	.10	17.67	18.06	-.39
2018	20.92	20.38	.54	18.28	18.34	-.06
2022	20.83	20.53	.30	18.16	18.63	-.47
2026		20.67			18.91	

Table 8: Men and Women Shot Put Distances Plus Best Fitted Model, Predicted Times And Residuals 1938 To 2022.Men RMSE=0.597m, Women RSME= 0.584m.

The discus data derived a cubic regression model as the best for both men and women discus athletes as displayed in table 9. The prediction for men indicated significant performance decline from 64.51m in 2022 to 62.32m in 2026. The model also predicts a performance plateau for males suggesting a performance limit may have been reached. The women discus athletes were predicted to display significant performance increases from 57.56m in 2022 to 63.81m in 2026. The model also predicts further performance increases beyond 2026. The RMSE values were for men 1.547m (error 2.4%) and women 3.014m (error 5.24%).

Year	Men Mean score (m)	Predicted distance(m) Cubic model Mean scores R^2=0.941, p<.001	Residual(m)	Women Mean score (m)	Predicted distance(m) Cubic model Mean scores R^2=0.775, p<.001	Residuals(m)
1938	41.49	40.27	1.22		40.09	
1950	42.66	46.34	-3.68		43.32	
1954	47.47	48.15	-.68	36.68	44.40	-7.72
1958	50.50	49.85	.65	43.74	45.48	-1.74
1962	51.32	51.45	-.13	45.19	46.56	-1.37
1966	54.27	52.94	1.33	47.70	47.64	.06
1970	56.32	54.32	2.00	49.38	48.72	.66
1974	56.33	55.59	.74	53.01	49.79	3.22
1978	56.24	56.76	-.52	56.05	50.87	5.18
1982	60.54	57.82	2.72	55.80	51.95	3.85
1986	57.21	58.78	-1.57	53.86	53.03	.83
1990	59.95	59.62	.33	54.02	54.11	-.09
1994	59.07	60.36	-1.29	56.78	55.19	1.59
1998	61.69	60.98	.71	58.88	56.26	2.62
2002	60.38	61.50	-1.12	57.96	57.34	.62
2006	61.63	61.91	-.28	59.62	58.42	1.20
2010	59.53	62.21	-2.68	58.92	59.50	-.58
2014	62.00	62.41	-.41	60.15	60.58	-.43
2018	63.10	62.49	.61	58.94	61.66	-2.72
2022	64.51	62.46	2.05	57.56	62.73	-5.17
2026		62.32			63.81	

Table 9: Men and Women Discus Distances Plus Best Fitted Model, Predicted Times And Residuals 1938 To 2022.Men RMSE=1.547m, Women RSME= 3.014m.

Table 10 represents the actual performances at the 2022 games and the predicted performances at the 2026 games for men and women across the seven events, as well as the direction of the performances changes. For men increased performances were predicted in 200m, 400m, high jump, long jump and decreased performance for 100m, shot put and discus. For women increased performances were predicted in 200m, 400m, shot put and discus and decreased in 100m, high jump and long jump.

	Actual	Predicted	Performance Change	Actual	Predicted	Performance Change
Event	Men 2022	Men 2026		Women 2022	Women 2026	
100m(s)	10.08	10.11	decrease	11.12	11.27	decrease
200m(s)	20.47	20.31	increase	22.84	22.26	increase
400m(s)	45.19	44.67	increase	51.04	50.64	increase
High Jump(m)	2.23	2.25	increase	1.92	1.89	decrease
Long Jump(m)	8.03	8.05	increase	6.85	6.58	decrease
Shot Put(m)	20.83	20.67	decrease	18.16	18.91	increase
Discus(m)	64.51	62.32	decrease	57.56	63.81	increase

Table 10: Summary Predicted Performances Based Averaged Performance of 1-6 Finalists In Each Event At The 2026 Commonwealth Games Compared To 2022 Actual Performance.

5. DISCUSSION

PREDICTING SPRINT, JUMP, AND THROWS PERFORMAMCE AT 2026 COMMONWEALTH GAMES

The Commonwealth Games data set from 1938 to 2022 derived some interesting results when compared to previous models based on Olympic data. The men's events in this research all the best fitted models were cubic functions, however they were consistent with the Heazlewood (2024) on the 2024 Paris Olympic study and the Heazlewood and Walsh (2015) study based on a 54 year data set from all years 1960 to 2014, where all men's events of 100m, 200m, 400m, high jump. Long jump, shot put and discus were described by the cubic function model. In this research predicting the men's performances in the 2026 Commonwealth Games the R^2 values were very significant from 0.708 for 400m to 0.941 for discus as indicated by table 4 in the results section. However, the results are inconsistent with previous research (Heazlewood and Lackey, 1996: Heazlewood, 2006a, 2006b), where the functions for the 100m was inverse, 400m was sigmoidal, long jump was cubic, long jump was compound, logistic, exponential and growth, and high jump was compound, logistic, exponential and growth. However, the latest research is based a larger set of data over more years of competition and now reflects more realistically the actual developmental trends in the men's 100m, 200m, 400m, high jump, long jump, shot put and discus. The concept is more data used the more representative the model.

The predictions for the events studied for the 2026 Commonwealth Games, Glasgow are presented in table 12 in absolute values and represent marginal increases in 200m (0.78%), 400m (1.15%) high jump (0.89%), long jump (0.23%) and marginal decreases in shot put and discus decreases in 100m (0.29%), shot put (0.77%) and discus (3.39%). All these predictions are plausible based on the high explained variance, very good fit between the cubic regression model and the data and realistic percentage changes in performance.

The women's results were in terms of the best function was the cubic regression model with high R^2 values from 0.679 for 200m to 0.920 for 100m. These results were consistent with the results from the Heazlewood and Walsh (2015) study where 100m, 200m, 400m, high jump, long jump, shot put and discus changes with time in years were best fitted with a cubic regression function. However, these results in terms of best fit function are only partially consistent with the 2024 Paris Olympic Games results where 100m, high jump, and discus best fits were for the cubic function, whereas 200m, 400m and shot put were sigmoidal and women's long jump was inverse. However, the recurring cubic regression function being the best fit across other studies cited evaluating the athletic events in this study indicates that developmental trends in these events conform to an underpinning factor or factors that constrain human progress in these events of human sprints, jumps and throws in both men and women.

Both the numerical data by games year, genders and events tables and presented graphical figure for the 100m indicate we approaching or have attained performance plateaus as static performances have been identified for the men and women 100m, men 200m, men and women high jump, men and women long jump and men discus. In some events moderate increases in performance are predicted for women 200m, men and women 400m, women shot and discus. In the men's discus a decrease in performance is predicted. These predicted performance are indicating limits of athletic performance have been attained as high performing international athletes, that is finalist at Olympic Games, Commonwealth Games and World Athletic Championships have optimal training, optimal nutrition, optimal technical skills for their events, optimal equipment and optimal competition venues in which they compete at international athletic competitions.

6. CONCLUSIONS

PREDICTING MEN AND WOMEN EVENT PERFORMANCE AT 2026 COMMONWEALTH GAMES

1. Men and women athletes derived a cubic regression function as the best fit based on data in the 100m, 200m, 400m, high jump, long jump, shot put and discus. In this context the development trends were common to both genders across these events, and which describe and quantify change over times.
2. The fit between the cubic models and the data for all the events was very good with high explained variance (R2 values), high statistical significance ($p<.001$) and low root mean square error (RMSE) for events with gender.
3. Plausible numerical predictions were derived based on the event performance trends from 1938 to 2022 Commonwealth Games to predict performances of the 1 to 6 finalists at the 2026 Commonwealth Games, Glasgow in the 100m, 200m, 400m, high jump, long jump, shot put and discus athletic events. The answer to these predictions will be known with certainty post Commonwealth Game 2026.

References

Garson, D.G. (2010) Curve estimation. Statnotes: Topics in Multivariate Analysis. Retrieved 1/12/2010 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Hair, J.E., Block, W., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) Multivariate Data Analysis. (6th Ed.). Pearson - Prentice Hall.

Heazlewood, I. & Lackey, G. (1996). The use of mathematical models to predict elite athletic performance at the Olympic Games. Conference proceedings 3rd Conference on Mathematics and Computers in Sport. Bond University (30 September - 2 October, 1996) Queensland.

Heazlewood, I. & Lackey, G. (1998). Mathematical models that predict performance decline in elite veteran athletes 30-90 years in the sprint distance and jump events. Conference proceedings 4th Conference on Mathematics and Computers in Sport. Bond University (13-15 July, 1998), Queensland. Pp.51-78.

Heazlewood, T. (2006a). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Conference Proceedings of the Eighth Australasian Conference on Mathematics and Computers in Sport. 3-5 July 2006, pp. 2-11. Coolangatta, Queensland, Australia.

Heazlewood, T. (2006b). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Journal of Sports Science and Medicine. 5, pp. 541-547.

Heazlewood, I. (2008) Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Oral presentation to Australian Statistical Society (NSW Branch).University of Sydney. 24/9/2008.

Heazlewood, I. (2010). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Model Predicting 2012 Performances. 2010 High Performance Sports Research Workshop, Australian Institute of Sport. Canberra, Australia 12-14 May, 2010. Invited Poster Presentation.

Heazlewood, I.T. & Walsh, J. (2011a). Mathematical models that predict 2012 London Olympic Games swimming and athletic performances. In Proc. 3rd International Conference on Mathematics in Sport. Pp.71-78. Salford, UK.

Heazlewood, I.T. & Walsh, J. (2011b). Mathematical Models that predict athletic performance in sprint and jump events for current and future world champions. In Proc. 3rd International Conference on Mathematics in Sport. Pp.64-70. Salford, UK.

Heazlewood, I. (2013a). The congruence between mathematical models predicting 2012 London Olympic Games swimming and athletic Performances and Actual Performance. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation

Heazlewood, I. (2013b). The congruence between mathematical models predicting 2011 Athletic World Championship performances and actual performance in the 100m, 400m, long jump and high jump. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation.

Heazlewood, I.T. & Walsh, J. (2014a). Mathematical models that predict athlete performance in the men's throwing events at the Olympic Games. Conference proceedings 12th Australian Conference on Mathematics and Computers in Sport. (25-27 June, 2014), Darwin, Australia. Pp. 48-53.

Heazlewood, I.T. & Walsh, J. (2014b). Mathematical models that predict athlete performance in the women's throwing events at the Olympic Games. Conference proceedings 12th Australian Conference on Mathematics and Computers in Sport. Bond University (25-27 June, 2014), Darwin, Australia. Pp. 54-59.

Heazlewood, I.T. & Walsh, J. (2015). Mathematical Models That Describe and Predict Performance Change in Men's And Women's Athletic Events: 1960-2014. Mathsport International, Loughborough University, UK 29 June – 1 July 2015. Oral presentation.

Heazlewood, I.T. (2024). The Use of Mathematical Models to Predict Athletic Performance in Men's And Women's 100m, 200m, 400m, Long Jump, High Jump, Shot Put and Discus at the 2024 Paris Olympic Games. The 17th Australasian Conference on Mathematics and Computers in Sport. Melbourne. 3-5 July 2024.Pp. 73-91.

Hopkins, W. (2000) Limits to performance. SportScience 4, Retrieved 3/4/2006 from http://www.sportsci.org/jour/0002/inbrief.html#limits. IAAF. (2008)

International Olympic Committee (2014). Olympic Results: Your source for event results and medalists from every Olympic Games. International Olympic Committee.

International Olympic Committee (2024). Olympic Results: Your source for event results and medalists from every Olympic Games. International Olympic Committee. https://olympics.com/en/olympic-games/olympic-results.

Lackey, G. & Heazlewood, I. (1998). Use of Mathematical Models to Predict Elite Swimming Performance. Conference proceedings 4th Conference on Mathematics and Computers in Sport. Bond University (13-15 July, 1998), Queensland. Pp. 79-110.

Norušis, M., SPSS for windows: Base systems user's guide release 6.0, (Chicago, SPSS Inc, 1993).

Prendergast, K., What do world running records tell us? Modern Athlete and Coach. 28 (1) 33 -36 (1990).

Walsh, J. & Heazlewood, I. (2014). Performance variability in the throwing events. Proceedings of the International Association of Computer Science in Sport Conference (IACSS2014). Ed. Assoc. Prof. Ian Heazlewood, Assoc. Prof. Anthony Bedford, Darwin, Australia, June 22-24. 2014. Pp. 109-115.

Wikipedia (2026). Athletics at the Commonwealth Games. https://en.wikipedia.org/wiki/Athletics_at_the_Commonwealth_Games

BAYESIAN NETWORK INFERENCE WITH UNCERTAINTY: MODELLING ATHLETE PROGRESSION

Madison Stibbe [a,b], Paul Wu [a]

[a] *Queensland University of Technology (QUT)*
[b] *Corresponding author: madison.stibbe@hdr.qut.edu.au*

Abstract
Athlete categorisation is a major contributing factor in a sporting body's ability to predict which current youth competitors may become the next generation of senior athletes representing Australia. This process is crucial, as the early identification of talent results in resources being conserved and allows for the support of athletes who have the greatest chance of being selected on a national team and winning medals in the future. The career progression of an athlete refers to the changes in an athlete's competitive performance over time, from local and junior competitions through to senior international representation. Whilst existing frameworks in the area of athlete development are able to categorise athletes based on their current performance level, they lack focus on the transitions between development stages and the influence of performances from one season to the next. This talk discusses a pilot framework for modelling athlete career progression centred around a Dynamic Bayesian Network (DBN), demonstrated through a case study in swimming athlete development. The proposed framework integrates expert knowledge with performance data to update DBN parameters while accounting for uncertainty. This involves generating samples from local Dirichlet-multinomial posterior distributions, alongside functional principal component analysis (FPCA) and Gaussian mixture model (GMM) techniques to estimate latent state probabilities and manage career trajectory variability. An algorithm was developed to perform DBN inference given observed evidence, enabling the model to be seasonally updated with new athlete performance data. The framework is designed to be easily generalisable through the use of relative performance metrics and athlete rankings and could be applied to other sports and contexts. The model could potentially play a role in talent identification and pathway planning by offering a data-driven, interpretable approach to managing athlete progression through the prediction of peak performance and variability.

Keywords: Athlete performance, talent identification, career trajectories, performance progression

Acknowledgements
With thanks to David Pyne and Lachlan Mitchell for their provision of data and insight.

References
Allen, S. V., Vandenbogaerde, T. J., & Hopkins, W. G. (2014). Career performance trajectories of Olympic swimmers: Benchmarks for talent development. *European Journal of Sport Science*, *14*(7), 643-651.
Barreiros, A., Côté, J., & Fonseca, A. M. (2012). From early to adult sport success: Analysing athletes' progression in national squads. *European Journal of Sport Science*, *14*(S1).
Born, D. P., Bjorklund, G., Lorentzen, J., Stoggl, T., & Romann, M. (2024). Specialize Early and Select Late: Performance Trajectories of World-Class Finalists and International- and National-Class Swimmers. *Int J Sports Physiol Perform*, *19*(2), 164-172.
Griffin, J. E., Hinoveanu, L. C., & Hopker, J. G. (2022). Bayesian modelling of elite sporting performance with large databases. *Journal of Quantitative Analysis in Sports*, *18*(4), 253-268.
Peterson, K. D., & Evans, L. C. (2019). Decision Support System for Mitigating Athletic Injuries. *International Journal of Computer Science in Sport*, *18*(1), 45-63.
Post, A. K., Koning, R. H., Visscher, C., & Elferink-Gemser, M. T. (2019). Multigenerational performance development of male and female top-elite swimmers–A global study of the 100 m freestyle event. *Scandinavian Journal of Medicine & Science in Sports*, *30*(3), 564-571.
Santos-Fernandez, E., Wu, P., & Mengersen, K. L. (2019). Bayesian statistics meets sports: a comprehensive review. *Journal of Quantitative Analysis in Sports*, *15*(4), 289-312.

Submitted to MATHSPORT 2026

THREE DECADES OF AI IN SPORTS-RELATED CONCUSSION RESEARCH: A BIBLIOMETRIC ANALYSIS (1996–2026)

Hanem Ellethy [a,b,e], Mostafa Kamalpour [a,c], Rania Shibl [a,d], Shahrzad Saremi [a], Mansoureh Mirzaei [a]
[a] *School of Science, Technology and Engineering, University of the Sunshine Coast, Brisbane, QLD 4502, Australia*
[b] *School of Information Technology and Electrical Engineering, University of Queensland, St Lucia, QLD 4070, Australia*
[c] *School of Information Systems, Queensland University of Technology, Brisbane, QLD 4000, Australia*
[d] *Faculty of Science and Engineering, Southern Cross University, Gold Coast, QLD, 4425, Australia*
[e] *Corresponding author: hellethy@usc.edu.au*

Abstract

Sports-related concussion, classified clinically as mild Traumatic Brain Injury (mTBI), is one of the most common yet least objectively diagnosed injuries in competitive athletics. An estimated 1.6 to 3.8 million sports concussions occur annually in the United States, and the Lancet Neurology Commission on TBI identifies mTBI as the dominant clinical form of brain injury, accounting for over 90% of cases, yet one for which evidence to guide diagnosis and management remains scarce. Artificial Intelligence (AI) and neuroimaging offer promising tools to address this gap; however, it remains unclear whether three decades of computational research have prioritised the high-incidence, 'invisible' mTBI or focused on more imaging-visible moderate-to-severe cases. This study systematically examines this prioritisation over three decades.

This study presents a bibliometric analysis of 1,950 publications retrieved from the Web of Science (1996–2026). Using VOSviewer for publication trend, keyword co-occurrence, and collaboration network analyses, we map the intellectual structure and thematic evolution of AI and neuroimaging research for TBI and concussion. Our analysis reveals a fourfold growth in annual research output over the past decade (63 publications in 2016 to 257 in 2025), with the United States leading (979 publications), followed by China (270) and the United Kingdom (169). Dominant themes include traumatic brain injury, machine learning, biomarkers, concussion, and deep learning. Critically, sports-specific concussion terminology occurs 4.4 times less frequently than severe-TBI clinical markers in the keyword landscape, despite its larger global burden.

These findings provide a roadmap for the computational sports science community to target the most underserved challenge in brain injury diagnosis: objective, imaging-based tools to support return-to-play decisions in athletic populations, particularly for mTBI. The analysis also highlights emerging AI techniques, such as explainable AI and federated learning, as promising avenues for future research.

Keywords: Sports-related concussion; mTBI; AI; ML; Neuroimaging; Bibliometric analysis; Clinical decision support; Diagnostic biomarkers; Return-to-play.

1. INTRODUCTION

The most prevalent form of traumatic brain injury is paradoxically the hardest to detect. Mild Traumatic Brain Injury (mTBI), commonly known as concussion, accounts for over 70% of all TBI cases in the United States and approximately 91% globally (Ellethy, Chandra, and Nasrallah 2021). Despite this enormous burden, standard neuroimaging, the clinical cornerstone of brain injury assessment, yields normal results in the vast majority of mTBI cases. This diagnostic invisibility is not incidental; it reflects a fundamental characteristic of mTBI, where injury occurs at the microstructural level, below the resolution threshold of conventional CT (Korley et al. 2016). In stark contrast, the same imaging tools reliably detect the haemorrhages, contusions, and midline shifts associated with moderate-to-severe TBI, making those cases computationally tractable in ways that mTBI is not (Anjum, Sakib, and Masudul Ahsan 2023; Asif et al. 2023).

The scale of this undetected burden is substantial. An estimated 3.9 million CT scans are performed annually in the United States to assess TBI patients, yet 91% return negative results and only 22% of TBI-diagnosed patients show positive CT findings (Chen et al. 2017; Maas et al. 2022). The Centers for Disease Control and Prevention (CDC) estimates that current surveillance systems capture only around one in every nine concussions nationally, indicating that the true national burden is considerably larger than reported figures suggest (U.S CENTERS FOR DISEASE CONTROL AND PREVENTION 2026). Sports and recreational activities account for 1.6 to 3.8 million of these concussions annually, with approximately 70% of sports-related TBI emergency department visits affecting children and adolescents aged 17 and under (Voss et al. 2015). In 2019, approximately 15% of US high school students reported a sports-related concussion in the preceding year (Gardner, Quarrie, and Iverson 2019).

The consequences of mismanaged concussion extend well beyond the immediate injury. Without objective diagnostic tools, return-to-play decisions in competitive sport rely on subjective symptom checklists, instruments prone to under-reporting, particularly in high-stakes athletic environments where athletes may conceal symptoms to remain in competition (Hobbs, Young, and Bailes 2016) . Premature return to play following unresolved concussion carries documented risks of second-impact syndrome, prolonged post-concussion syndrome, and, with repeated injury, chronic traumatic encephalopathy (CTE) (Cantu and Uretsky 2021). These outcomes are not merely clinical concerns; they are a sports science and public policy problem demanding objective, data-driven solutions. The Lancet Neurology Commission on TBI frames this gap explicitly. TBI is identified as the highest-incidence neurological disorder globally, yet the Commission notes that more than 90% of patients presenting to hospital with TBI have so-called mTBI, but evidence to inform treatment of patients with mild TBI is scarce, calling for increased public health interest and a dedicated research focus on mTBI. The Commission further highlights that around 50% of patients with mTBI do not recover to pre-injury health and wellbeing by six months and calls for research to identify those at highest risk of incomplete recovery, exactly the kind of stratification task for which Artificial intelligence (AI) and neuroimaging methods are suited (Maas et al. 2022; Van Pelt et al. 2021).

AI and machine learning (ML) have transformed medical imaging analysis over the past three decades, offering the computational capacity to detect patterns invisible to human observers (Mehta et al. 2022). Within TBI research specifically, ML approaches have demonstrated remarkable performance in haemorrhage segmentation, lesion classification, and outcome prediction for moderate-to-severe injury (Lin and Yuh 2022). However, prior work has identified a critical asymmetry: the overwhelming majority of ML and neuroimaging research has focused on moderate-to-severe TBI, where structural abnormalities provide clear computational targets, while mTBI, despite representing over 90% of all TBI cases globally, has received disproportionately little attention. A comprehensive review of CT-based ML studies by the present authors found that only one of 26 studies specifically targeted mTBI (Ellethy 2024). Because sports-related concussion is a subset of mTBI, this underrepresentation necessarily extends to the sports context; however, the scale and structure of that extension across three decades of research has not previously been quantified. Answering whether this pattern of neglect has persisted or shifted across three decades of research activity requires a method suited to mapping fields at scale rather than evaluating individual studies, which is precisely what bibliometric analysis offers (Donthu et al. 2021).

This raises a fundamental question: does three decades of AI and neuroimaging research reflect the actual clinical burden of brain injury, or has the field systematically followed the path of least computational resistance?

The present study focuses specifically on imaging-based AI research in TBI and concussion, the intersection where diagnostic invisibility creates the greatest computational challenge. The search strategy was therefore constrained to studies combining AI or machine learning methods with neuroimaging modalities (CT, MRI, fMRI, DTI, EEG, or PET) in the context of TBI or concussion. This study addresses this question through a systematic bibliometric analysis of 1,950 scholarly publications retrieved from the Web of Science database spanning 1996 to 2026. A bibliometric approach is warranted because the underlying question is not whether individual AI studies perform well on their chosen datasets, but whether the field as a whole has directed its attention toward the most clinically significant problem. Bibliometric analysis uniquely captures the scale, geographic distribution, collaboration structure, and thematic evolution of research activity across decades, patterns that individual study reviews or meta-analyses cannot reveal (Donthu et al. 2021; Kamalpour, Nili, and Yigitcanlar 2025). Using publication trend analysis, keyword co-occurrence mapping, and collaboration network analysis, we map the intellectual structure and thematic evolution of AI-driven TBI and concussion research. Our analysis is guided by four research questions:

- **RQ1:** What are the publication trends in AI and neuroimaging research for TBI and concussion over 1996–2026?
- **RQ2:** Which countries, institutions, and authors lead this research?
- **RQ3:** What are the dominant research themes and emerging computational approaches?
- **RQ4:** Where are the critical gaps, particularly regarding sports-related mTBI and what directions should future research prioritise?

By mapping where three decades of research has, and critically, has not, been directed, this study provides a roadmap for the computational sports science community to target the most underserved and clinically significant challenge in brain injury diagnosis.

2. RELATED WORK

While the application of AI and neuroimaging in brain injury has garnered significant academic interest, a review of existing bibliometric literature reveals a highly fragmented research landscape. Previous bibliometric analyses

have generally mapped these domains in isolation, either focusing broadly on AI in generalized trauma (Wang, Zhang, and Zhao 2024), tracking two decades of general TBI research (Ye et al. 2024), or mapping general publication trends without isolating specific diagnostic modalities (Long et al. 2024). As noted by Long et al. (2024), previous bibliometric studies in this field have typically focused too broadly on disease classifications or limited themselves to specific, isolated populations (Long et al. 2024).

By analysing the existing reviews, three distinct gaps emerge which the current study addresses:

The Misalignment of AI and Epidemiological Burden Recent bibliometric analyses confirm that mild traumatic brain injury (mTBI) accounts for the vast majority, between 70% and 90%, of all global TBI cases (Ramlatchan et al. 2025). However, the literature shows that AI and computational research has struggled to align with this reality. While recent topic modelling suggests a gradual thematic shift in the literature from severe head-injury toward mTBI, the primary epidemiological focus has shifted toward demographic changes, such as the rising incidence of TBI due to falls among the elderly (Ye et al. 2024). Consequently, while the broader impact of AI in trauma is recognized as a critical future frontier (Wang, Zhang, and Zhao 2024), the massive, specific clinical burden of sports-related concussion remains a highly under-analysed niche within computational diagnostics.

The Over-Reliance on Molecular Diagnostics over Computational Imaging When reviewing the specific literature on mTBI diagnostics, existing bibliometrics reveal a heavy bias toward molecular and fluid biomarkers. A recent analysis spanning two decades of mTBI diagnostics found that molecular markers are currently the most studied diagnostic tool, accounting for 28.4% of all publications in the field (Liu et al. 2023). While fluid biomarkers represent a growing trend, they often require invasive sampling or are limited by the blood-brain barrier. In contrast, neuroimaging is an integral, non-invasive component of neurological practice (Ramlatchan et al. 2025), yet its intersection with AI for the specific detection of sports-related mTBI remains critically under-mapped.

The Need for a Three-Way Intersectional Analysis Although studies have analysed the top-cited publications in mTBI (Shi et al. 2021) and the broad global landscape of neuroimaging in TBI (Ramlatchan et al. 2025), no study to date has isolated the specific technological triad required for modern athletic diagnostics. The current bibliometric literature confirms that while clinical studies have historically predominated mTBI research, investigations into injury mechanisms and early diagnostic tools have lagged behind (Shi et al. 2021).

Novelty of this Study: To our knowledge, this paper is the first to systematically map the three-way intersection of AI, neuroimaging, and sports-related mTBI over a 30-year period. By moving beyond general TBI or isolated molecular biomarker reviews, this study uniquely exposes how computational research has historically favoured severe, radiologically obvious traumas over the complex, microstructural anomalies required to diagnose "invisible" sports-related concussions.

3. METHODOLOGY

DATA SOURCE AND SEARCH STRATEGY

To systematically evaluate the literature at the intersection of AI, neuroimaging, and mTBI, a comprehensive search was conducted using the Web of Science Core Collection database (Science n.d.). The search was executed on April 9, 2026, explicitly targeting literature published over a three-decade span between 1996 and 2026.

The search strategy was strictly constrained to identify studies that combined computational and ML methods with diagnostic neuroimaging modalities in the specific clinical contexts of mTBI or concussion. To capture this, the exact search string utilized the following Boolean operators:

(concussion OR mTBI OR "mild traumatic brain injury" OR TBI) AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR AI OR ML OR "neural network" OR CNN) AND (neuroimaging OR imaging OR MRI OR fMRI OR DTI OR EEG OR CT OR brain).

This primary search strategy yielded a final, relevant dataset of 1,950 scholarly publications. The Web of Science Core Collection inherently deduplicates records during a single search query, returning unique publication records. The exported dataset of 1,950 records was nevertheless inspected for any residual duplicates based on Digital Object Identifier (DOI) and title matching prior to analysis; no duplicates were identified.

DATA EXTRACTION AND BIBLIOMETRIC ANALYSIS

Following data retrieval, the full bibliographic records of the 1,950 documents, including authorship, affiliations, publication year, citations, and keywords, were extracted for analysis. The bibliometric mapping and network visualization were conducted using VOSviewer software (Van Eck and Waltman 2007). The analytical approach was designed to systematically address the study's four research questions. VOSviewer was utilized to conduct publication trend analysis, map keyword co-occurrence networks to reveal dominant and emerging research

themes and generate collaboration networks to identify the geographic and institutional leadership driving AI-based neuroimaging research in brain injuries.

4. RESULTS AND DISCUSSION

Publication trends in AI, neuroimaging, and TBI (RQ1): An analysis of the publication trends between 1996 and 2026 reveals a dramatic acceleration in research applying AI and machine learning to TBI and concussion. A total of 1,950 papers were identified, demonstrating a clear exponential growth curve over the last decade, as visualized in Figure 1. In 2016, the field produced 63 publications. By 2020, this number had more than doubled to 143 publications. This upward trajectory continued, reaching 199 publications in 2024 and peaking at 257 in 2025, representing a fourfold growth in annual research output over the past decade. This surge underscores the rapidly growing recognition of AI's utility in addressing complex diagnostic challenges in brain injuries.

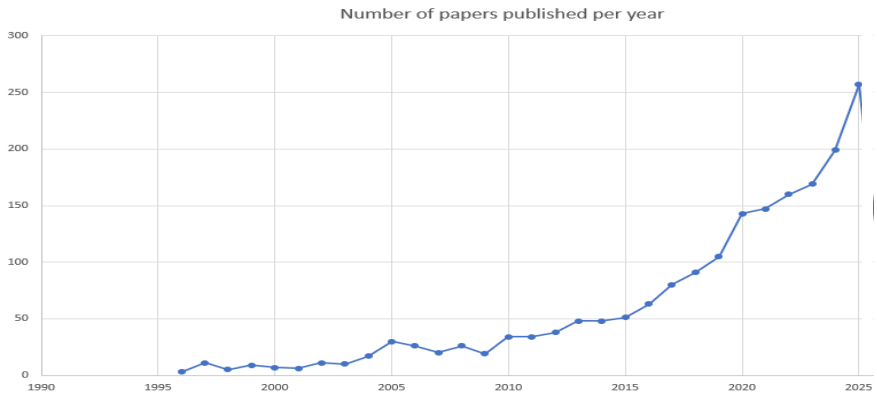


Figure 1. Annual publication output in AI and neuroimaging research for TBI and concussion (1996–2026), based on Web of Science Core Collection records (n = 1,950) retrieved on 9 April 2026. The 2026 value (82 papers as of April) is partial-year only. Annual output grew fourfold over the past decade, from 63 publications in 2016 to 257 in 2025.

Leading countries, institutions, and authors (RQ2) : The geographical and institutional distribution of research at the intersection of AI and neuroimaging is highly concentrated, revealing a field anchored by a few "super-hubs" but characterized by evolving regional clusters.

Country-level collaboration patterns: As depicted in Figure 2, the United States overwhelmingly dominates the field (979 documents; 27,112 citations). Beyond sheer volume, the co-authorship network reveals that the USA serves as the **global bridge**; it is the only node with significant connectivity to every other major cluster.

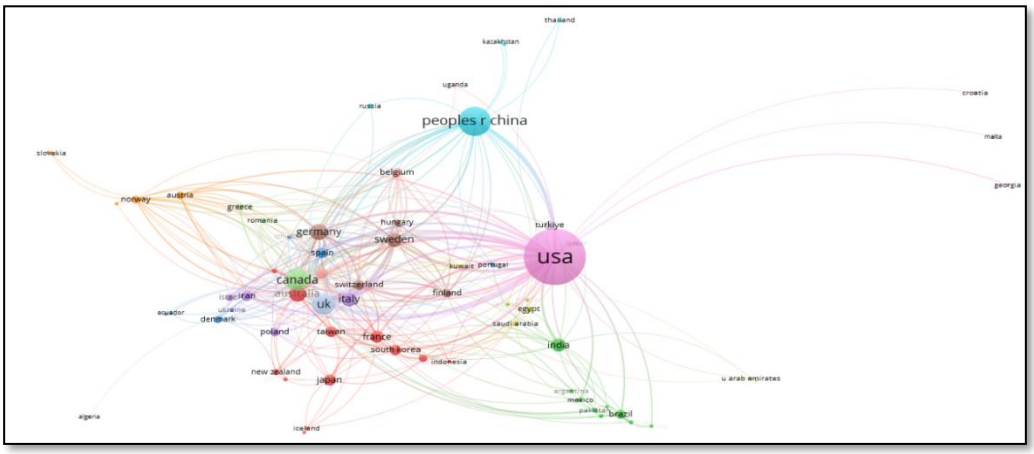


Figure 2. Country co-authorship network for AI and neuroimaging research in TBI and concussion (1996–2026), generated using VOSviewer. Node size reflects publication count; line thickness reflects co-authorship strength; colours indicate cluster membership

The network architecture exposes two distinct patterns:

- **The Transatlantic-Commonwealth Axis:** A dense, highly integrated core consisting of the USA, UK, Canada, Germany, and Italy. This suggests that the technical integration of AI into neuroimaging pipelines is heavily driven by established Western consortia like *TRACK-TBI* and *CENTER-TBI*.
- **The Emerging Asian Cluster:** While China ranks second in publication volume (270 documents), its network position is distinct. It serves as a regional hub for Asian nations (e.g., South Korea, Thailand) but maintains fewer direct "co-authorship edges" to the European core compared to the USA.
- **The Australasian Position:** Australia ranks sixth globally but sits at the periphery of the main Western cluster. This indicates that while Australian research is productive, it is more "insular" compared to the Canadian or British nodes, representing an opportunity for deeper international integration in sports-related concussion AI.
- **Institutional impact and productivity:** The institutional landscape, illustrated in Figure 3 and Figure 4, validates the network "hubs" seen in the mapping. The University of Pittsburgh leads globally in both volume (92 publications) and influence (50 citations per document). This dual-dominance explains its central, high-degree position in institutional network maps.

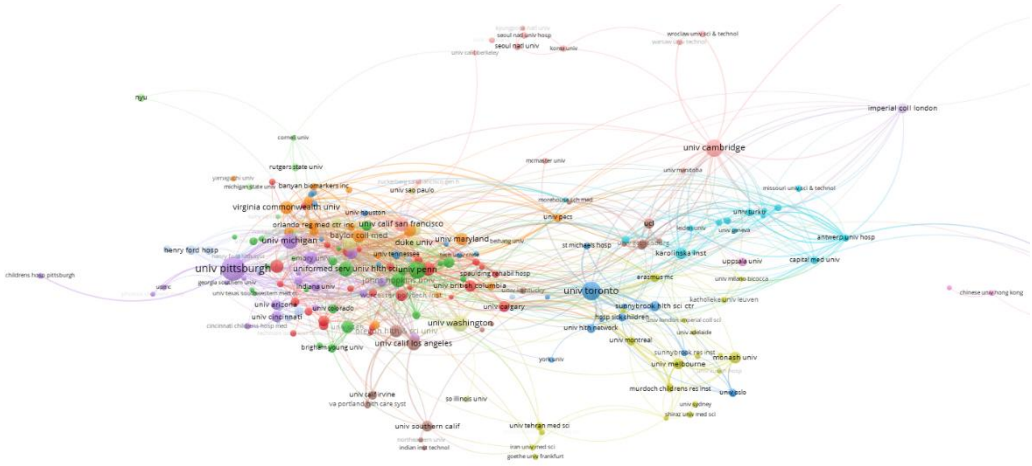


Figure 3. Institutional co-authorship network for AI and neuroimaging research in sports-related concussion (1996–2026). Node size is proportional to the total number of publications; line thickness represents the strength of the collaborative link between institutions; and colours indicate distinct modularity clusters.

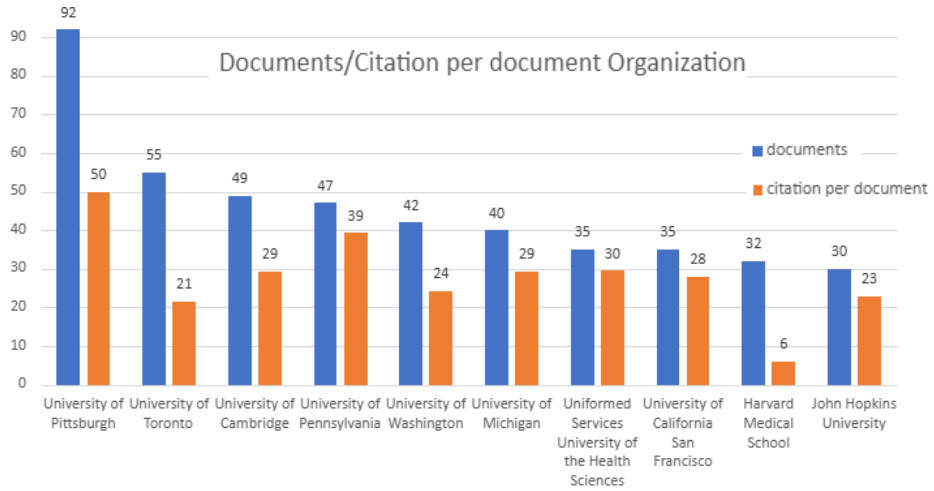


Figure 4. Top ten contributing institutions ranked by publication count, with corresponding citations per document. The University of Pittsburgh leads in both volume (92 publications) and impact (50 citations per document)

In contrast, a comparison between the **University of Toronto** and **University of Pittsburgh** reveals different collaborative roles:

- **Toronto** acts as a "Volume Engine" (55 documents) but with lower relative impact (21 citations/doc).
- **Pittsburgh** acts as an "Influence Anchor," providing the foundational research that others cite.

- **Harvard Medical School** presents a unique outlier; despite its prestige, its AI-neuroimaging concussion documents show a lower citation rate (6 per doc), suggesting its entry into this specific niche may be more recent or focused on emerging, less-cited pilot methodologies.
- **Author leadership and knowledge cliquishness:** Analysis of individual contributions, Figures 5 & 6 identify a tension between productivity and impact. Anthony E. Kline is the most prolific author (29 documents), anchoring a significant research cluster focused on experimental TBI. However, the intellectual "weight" of the clinical-AI intersection is carried by authors like Linda Papa. Despite having fewer publications (15), her staggering impact of 176 citations per document makes her a "Super-Node" in the network.

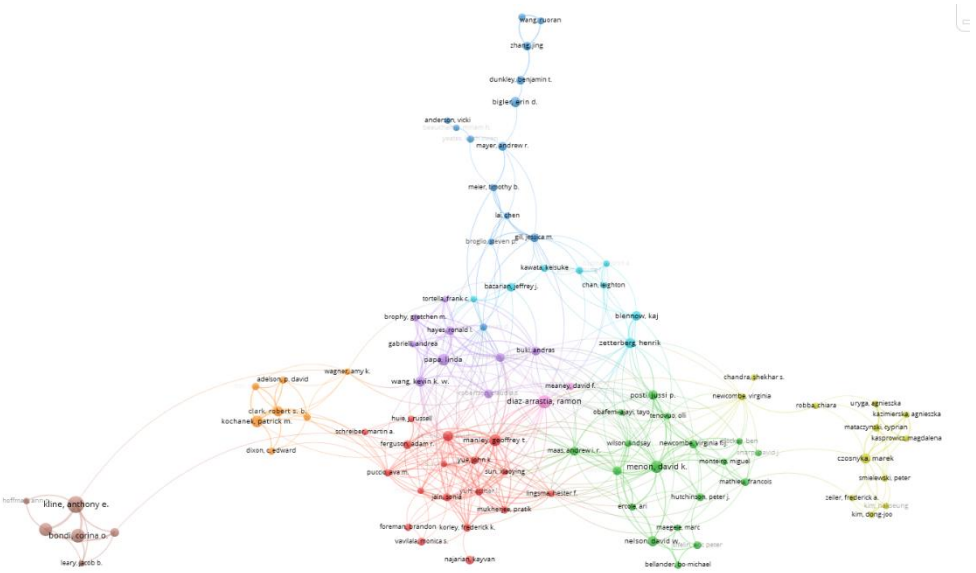


Figure 5. Author co-authorship network illustrating the intellectual landscape of the field (1996–2026).

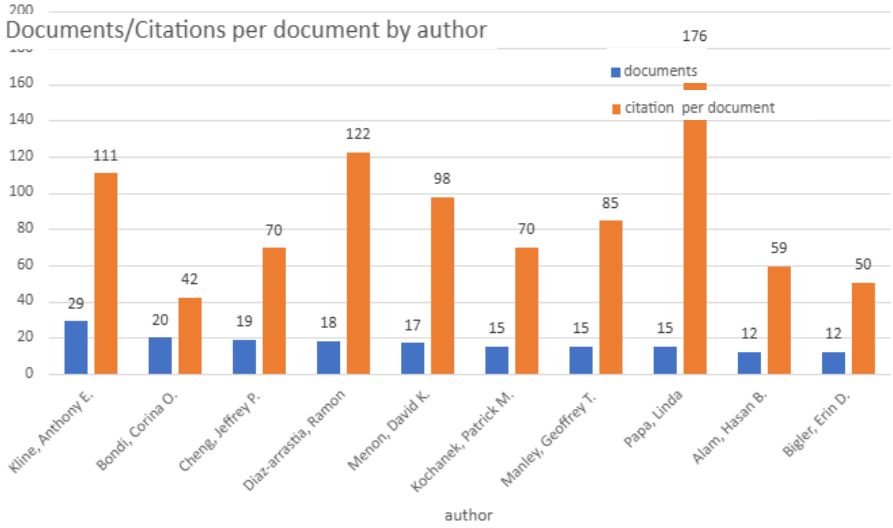


Figure 6. Top ten most prolific authors ranked by publication count, with corresponding citations per document.

Visual mapping of these authors reveals that the field is organized into "Tightly-Knit Cliques." High-impact authors like Ramon Diaz-Arrastia (122 citations/doc) and David K. Menon (98 citations/doc) often appear in the same clusters. This indicates that the transition of AI from theoretical models to clinical neuroimaging reality is being driven by a small, highly collaborative elite rather than a broad, decentralized group of researchers.

To provide a structured overview of the multi-level collaboration architecture, Table 1 summarizes the dominant patterns and their strategic implications for the field.

Analysis Level	Identified Collaboration Pattern	Strategic Interpretation
National	Transatlantic-Commonwealth Dominance	High level of policy and protocol alignment between US, UK, Canada, and Australia.
Institutional	The "Toronto Bridge" Model	The University of Toronto acts as a critical intermediary linking North American and European research clusters.
Author	Multi-modal Integration (AI + Biomarkers)	Leadership nodes (e.g., Zetterberg/Menon) indicate a shift from pure imaging to multi-modal AI diagnostics involving fluid biomarkers.
Regional	Australasian Peripheral Positioning	Despite high output, Australian nodes show relative isolation, suggesting a need for more formal integration into Euro-American AI consortia.

Table 1. Analysis of collaboration architecture across national, institutional, and author levels.

Dominant research themes and emerging computational approaches (RQ3): Keyword co-occurrence analysis exposes the intellectual structure of the field. Unsurprisingly, "traumatic brain injury" is the central node (1,098 occurrences). The dominant computational approaches are highly visible, with "machine learning" (289 occurrences), "deep learning" (80 occurrences), and "artificial intelligence" (73 occurrences) acting as primary methodological drivers.

To move beyond individual keywords and uncover the true thematic weighting of the field, keywords were grouped into research-theme clusters as showed in Table 2. This grouping reveals a pronounced imbalance: clinical markers of moderate-to-severe TBI (mortality, intracranial pressure, haemorrhage) generate 501 keyword occurrences across 19 distinct terms, while sports-specific terminology (football, athletic, sports-related concussion, etc.) yields only 114 occurrences across 10 distinct terms, a 4.4-fold imbalance.

Research theme	Distinct keywords	Total occurrences
AI/ML methods	12	572
Severe TBI clinical markers	19	501
mTBI / concussion (all variants)	13	423
Paediatric populations	7	168
Sports-specific terms	10	114
Return-to-play	0	0

Table 2. Thematic grouping of keyword occurrences in 1,950 AI-driven TBI publications (1996–2026)

Most striking is the complete absence of "return-to-play" as an indexed keyword in any of the 1,950 publications analysed. Despite return-to-play representing the most consequential clinical decision in sports concussion management, no publication in three decades of AI-driven TBI research indexes this term.

Neuroimaging modality distribution: Given the study's focus on imaging-based AI research, a dedicated analysis of neuroimaging modality keywords was conducted and reported in Table 3. Twenty-four distinct imaging-related keywords were identified, totalling 473 occurrences across the dataset.

Modality	Distinct keywords	Total occurrences
fMRI / functional connectivity	7	132
Structural MRI	2	89
Computed tomography (CT)	3	80
DTI / white matter	5	62
EEG	2	50
Generic "neuroimaging"	1	31
MEG	2	19
Spectroscopy (NIRS)	2	10

Table 3. Distribution of neuroimaging modality keywords across 1,950 publications

Three observations emerge from this distribution. First, functional MRI and functional connectivity dominate the indexed imaging vocabulary (132 occurrences), followed by structural MRI (89) and computed tomography (80). When MRI-based modalities are combined (structural MRI, fMRI, DTI), they total 283 occurrences, approximately 3.5 times the CT count. This pattern is notable because CT is the clinical standard for acute sports concussion presentation in emergency departments (Ellethy, Chandra, and Nasrallah 2022); yet AI research has disproportionately focused on MRI-based methods, likely reflecting the greater availability of research-grade MRI datasets from initiatives such as TRACK-TBI (Nelson et al. 2021) and CENTER-TBI (Zeiler et al. 2020).

Second, DTI and white matter analysis, the imaging modality best suited to detecting the microstructural axonal injury characteristic of mTBI (Jang and Cho 2022), appears in only 62 keyword occurrences across five terms. This represents less than half the attention given to fMRI, despite DTI's particular relevance to concussion pathophysiology.

Third, PET imaging is entirely absent from the indexed keyword landscape. This is a significant gap, as PET-based tau imaging is a key research tool for detecting CTE in retired athletes with a history of repeated concussion (Ayubcha et al. 2021).

Critical gaps: the underrepresentation of sports-related mTBI (RQ4): The keyword analysis confirms a severe misalignment between research priorities and the clinical epidemiology of TBI. While the general term "concussion" appears 171 times, more specific mTBI terminology is considerably less prevalent: "mild traumatic brain injury" (104 occurrences), "mtbi" (36), "postconcussion syndrome" (20), and "sports-related concussion" (19). The sports context specifically is strikingly marginal: sport-related keywords (sport, football, soccer, athletic, and variants) collectively occur 114 times across the entire dataset.

Five findings crystallise the research-to-burden mismatch:

1. **A 4.4-fold imbalance** between severe-TBI clinical markers (501 occurrences) and sports-related terminology (114 occurrences).
2. **Return-to-play has zero indexed occurrences** across 1,950 publications, despite being the central clinical decision in sports concussion management.
3. **Sports-specific terms constitute approximately 1.3% of the indexed keyword vocabulary**, strikingly low relative to the epidemiological burden of sports concussion (1.6 to 3.8 million cases annually in the United States alone).
4. **CT research is underrepresented relative to clinical practice** , MRI-based research outnumbers CT-based research 3.5-fold, despite CT being the standard imaging modality in acute athletic presentation.
5. **PET imaging is absent** from the indexed landscape, representing a critical gap for CTE research in retired athletes.

These findings provide a quantitative answer to the fundamental question posed in this study: three decades of AI and neuroimaging research have not reflected the actual clinical burden of brain injury. Instead, the field has largely followed the path of least computational resistance, focusing on moderate-to-severe injuries with clear radiological findings rather than the microstructural anomalies characteristic of sports-related concussion.

Limitations: This analysis is based on author-indexed keywords and Keywords Plus terms extracted from the Web of Science database. Keyword-based bibliometrics reflect how authors and indexing systems label their work, which may not fully capture the content of every study, papers using CT imaging, for instance, may be indexed only under "traumatic brain injury" without an explicit imaging modality tag. However, the relative magnitude of the patterns reported here , including the 4.4-fold imbalance between severe-TBI and sports-related terminology and the 3.5-fold difference between MRI-based and CT-based research focus, exceeds the range attributable to indexing variation. Future full-text analyses could complement these findings.

5. CONCLUSION AND FUTURE DIRECTIONS

This bibliometric analysis of 1,950 publications from 1996 to 2026 confirms a critical asymmetry in AI and neuroimaging research. While overall research output in AI-driven brain injury assessment is growing exponentially, led predominantly by institutions in the United States, the focus remains disproportionately skewed toward moderate-to-severe TBI.

The high-incidence, "invisible" epidemic of sports-related concussion remains vastly underrepresented in the computational literature. Given that premature return-to-play can lead to devastating long-term consequences like chronic traumatic encephalopathy (CTE), there is an urgent need to redirect AI resources.

To translate these findings into action, we propose a three-stage research agenda for the computational sports science community:

Near-term (1–2 years): Develop and benchmark AI-based decision support tools using existing CT and DTI datasets specifically curated for sports-related concussion populations, with explainability frameworks designed for emergency department deployment.
Mid-term (3–5 years): Establish federated learning consortia across Australasian sports medicine centres to aggregate scarce, high-quality multimodal athletic concussion datasets, addressing both the data sparsity and privacy concerns that currently constrain the field.
Long-term (5+ years): Develop validated, real-time, multimodal AI tools that integrate imaging, biomarker, cognitive, and wearable-sensor data to support objective, evidence-based return-to-play protocols at all levels of competitive sport. Only by shifting computational attention to the hardest-to-detect injuries can the field meaningfully address the world's most prevalent neurological disorder

REFERENCES

Anjum, Nipa, Abu Noman Md Sakib, and Sk Md Masudul Ahsan. 2023. "Classification of Brain Hemorrhage Using Deep Learning from CT Scan Images." In *Proceedings of International Conference on Information and Communication Technology for Development: ICICTD 2022*, Springer, 181–93.

Asif, Muhammad, Munam Ali Shah, Hasan Ali Khattak, Shafaq Mussadiq, Ejaz Ahmed, Emad Abouel Nasr, and Hafiz Tayyab Rauf. 2023. "Intracranial Hemorrhage Detection Using Parallel Deep Convolutional Models and Boosting Mechanism." *Diagnostics* 13(4): 652.

Ayubcha, Cyrus, Mateen Moghbel, Austin J Borja, Andrew Newberg, Thomas J Werner, Abass Alavi, and Mona-Elisabeth Revheim. 2021. "Tau Imaging in Head Injury." *PET clinics* 16(2): 249–60.

Cantu, Robert C, and Madeline Uretsky. 2021. "Consequences of Ignorance and Arrogance for Mismanagement of Sports-Related Concussions: Short- and Long-Term Complications BT - Concussions in Athletics: From Brain to Behavior." In eds. Semyon M Slobounov and Wayne J Sebastianelli. Cham: Springer International Publishing, 3–17. doi:10.1007/978-3-030-75564-5_1.

Chen, Cheng, Junxin Shi, Rachel M. Stanley, Eric A. Sribnick, Jonathan I. Groner, and Henry Xiang. 2017. "U.S. Trends of ED Visits for Pediatric Traumatic Brain Injuries: Implications for Clinical Trials." *International Journal of Environmental Research and Public Health*. doi:10.3390/ijerph14040414.

Donthu, Naveen, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, and Weng Marc Lim. 2021. "How to Conduct a Bibliometric Analysis: An Overview and Guidelines." *Journal of business research* 133: 285–96.

Van Eck, Nees Jan, and Ludo Waltman. 2007. "VOS: A New Method for Visualizing Similarities between Objects." In *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the Gesellschaft Für Klassifikation EV, Freie Universität Berlin, March 8–10, 2006*, Springer, 299–306.

Ellethy, Hanem. 2024. "Deep Learning Strategies for Enhanced MTBI Diagnosis Using Clinical and CT Data."

Ellethy, Hanem, Shekhar S. Chandra, and Fatima A. Nasrallah. 2021. "The Detection of Mild Traumatic Brain Injury in Paediatrics Using Artificial Neural Networks." *Computers in Biology and Medicine* 135: 104614. doi:10.1016/J.COMPBIOMED.2021.104614.

Ellethy, Hanem, Shekhar S Chandra, and Fatima A Nasrallah. 2022. "Deep Neural Networks Predict the Need for CT in Pediatric Mild Traumatic Brain Injury: A Corroboration of the PECARN Rule." *Journal of the American College of Radiology* 19(6): 769–78. doi:https://doi.org/10.1016/j.jacr.2022.02.024.

Gardner, Andrew J, Kenneth L Quarrie, and Grant L Iverson. 2019. "The Epidemiology of Sport-Related Concussion: What the Rehabilitation Clinician Needs to Know." *journal of orthopaedic & sports physical therapy* 49(11): 768–78.

Hobbs, Jonathan G, Jacob S Young, and Julian E Bailes. 2016. "Sports-Related Concussions: Diagnosis, Complications, and Current Management Strategies." *Neurosurgical focus* 40(4): E5.

Jang, Sung-Ho, and Min-Jye Cho. 2022. "Role of Diffusion Tensor Imaging in the Diagnosis of Traumatic Axonal Injury in Individual Patients with a Concussion or Mild Traumatic Brain Injury: A Mini-Review." *Diagnostics* 12(7): 1580.

Kamalpour, Mostafa, Alireza Nili, and Tan Yigitcanlar. 2025. "Assessing Research Impact: A Framework and Implications for Management Information Systems Research."

Korley, Frederick K., Gabor D. Kelen, Courtney M. Jones, and Ramon Diaz-Arrastia. 2016. "Emergency Department Evaluation of Traumatic Brain Injury in the United States, 2009-2010." *Journal of Head Trauma Rehabilitation*. doi:10.1097/HTR.0000000000000187.

Lin, Emily, and Esther L Yuh. 2022. "Computational Approaches for Acute Traumatic Brain Injury Image Recognition." *Frontiers in neurology* 13: 791816.

Liu, Yishu, Chudong Wang, Zhiyun Pi, Tingting Wang, Changquan Zhang, and Jifeng Cai. 2023. "Research on the Potential Biomarkers of Mild Traumatic Brain Injury: A Systematic Review and Bibliometric Analysis." *Molecular neurobiology* 60(8): 4442–49.

Long, Yan-rui, Kai Zhao, Fu-chi Zhang, Yu Li, Jun-Wen Wang, Hong-quan Niu, and Jin Lei. 2024. "Trends and Hotspots in Research of Traumatic Brain Injury from 2000 to 2022: A Bibliometric Study." *Neurochemistry international* 172: 105646.

Maas, Andrew I R, David K Menon, Geoffrey T Manley, Mathew Abrams, Cecilia Åkerlund, Nada Andelic, Marcel Aries, et al. 2022. "Traumatic Brain Injury: Progress and Challenges in Prevention, Clinical Care, and Research." *The Lancet Neurology* 21(11): 1004–60.

Mehta, Ojas, Zhibin Liao, Mark Jenkinson, Gustavo Carneiro, and Johan Verjans. 2022. "Machine Learning in Medical

Imaging – Clinical Applications and Challenges in Computer Vision BT - Artificial Intelligence in Medicine: Applications, Limitations and Future Directions." In eds. Manda Raz, Tam C Nguyen, and Erwin Loh. Singapore: Springer Nature Singapore, 79–99. doi:10.1007/978-981-19-1223-8_4.

Nelson, Lindsay D, Mark D Kramer, Keanan J Joyner, Christopher J Patrick, Murray B Stein, Nancy Temkin, Harvey S Levin, et al. 2021. "Relationship between Transdiagnostic Dimensions of Psychopathology and Traumatic Brain Injury (TBI): A TRACK-TBI Study." *Journal of abnormal psychology* 130(5): 423.

Van Pelt, Kathryn L, Tim Puetz, Jennylee Swallow, Andrew P Lapointe, and Steven P Broglio. 2021. "Data-Driven Risk Classification of Concussion Rates: A Systematic Review and Meta-Analysis." *Sports Medicine* 51(6): 1227–44. doi:10.1007/s40279-021-01428-7.

Ramlatchan, Sidhartha R, Benjamin Colaco Jamal, Latha Ganti, Samiya Natesan-Torres, and Sreeja Natesan. 2025. "Neuroimaging in Traumatic Brain Injury: A Bibliometric Analysis." *International Journal of Emergency Medicine* 18(1): 168.

Science, Web of. "Smart Search." https://www.webofscience.com/wos/woscc/summary/6298333f-c2cc-4e6c-bb07-1a9f8ffcca2e-01ab7ddbb2/relevance/1 (April 9, 2026).

Shi, Jian, Xianping Du, María José Cavagnaro, and Na Li. 2021. "A Bibliometric Analysis and Visualization of the Top-Cited Publications in Mild Traumatic Brain Injury." *Frontiers in neurology* 12: 687796.

U.S CENTERS FOR DISEASE CONTROL AND PREVENTION. 2026. "National Concussion Surveillance System." https://www.cdc.gov/traumatic-brain-injury/programs/index.html.

Voss, Jameson D, Joseph Connolly, Karen A Schwab, and Ann I Scher. 2015. "Update on the Epidemiology of Concussion/Mild Traumatic Brain Injury." *Current Pain and Headache Reports* 19(7): 32. doi:10.1007/s11916-015-0506-z.

Wang, Chun, Mengzhou Zhang, and Dong Zhao. 2024. "Mapping the Impact of Artificial Intelligence on Trauma Research via Scientometric Analysis." *Journal of Forensic Science and Medicine* 10(2): 133–47.

Ye, Ziyin, Zhi Li, Shiyu Zhong, Qichen Xing, Kunhang Li, Weichen Sheng, Xin Shi, and Yijun Bao. 2024. "The Recent Two Decades of Traumatic Brain Injury: A Bibliometric Analysis and Systematic Review." *International Journal of Surgery* 110(6): 3745–59.

Zeiler, Frederick A, Ari Ercole, Erta Beqiri, Manuel Cabeleira, Eric P Thelin, Nino Stocchetti, Ewout W Steyerberg, et al. 2020. "Association between Cerebrovascular Reactivity Monitoring and Mortality Is Preserved When Adjusting for Baseline Admission Characteristics in Adult Traumatic Brain Injury: A CENTER-TBI Study." *Journal of Neurotrauma* 37(10): 1233–41.

Quantitative Modeling of Mental Health Improvement Through Aerial Sports Participation and Nutrition Awareness

Shaden Khaled Aldakheel[a], Arwa Hamad M. Alogaili[a], Shahrzad Saremi[b], Mansooreh Mirzaei[b], Maryam Nooraei Abadeh[c], Ali Sakhaei[d], Rania Shibl[e], Anthony Bedford[b], Erica Mealy[b]

[a] School of Art and Design, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia
[b] School of Science, Technology and Engineering, University of the Sunshine Coast, Brisbane, Australia
[c] Department of Computer Engineering, Arv. C., Islamic Azad University, Abadan, Iran
[d] Department of Statistics, Payame Noor University (PNU), Tehran, Iran
[e] Faculty of Science and Engineering, Southern Cross University, Gold Coast, Australia

ABSTRACT

This research develops a quantitative model of mental health improvement associated with aerial sports participation and nutrition awareness, with implications for rehabilitation environment design. A mixed-method approach was employed, combining survey-based quantitative analysis with qualitative interviews. Key psychological and behavioural constructs were operationalized as measurable variables using ordinal and categorical scales. Survey responses were numerically encoded to enable statistical modeling. A composite indicator, termed the Mental Health Improvement Index (MHII), was constructed by aggregating variables related to stress reduction, perceived relaxation, and emotional recovery. Using 108 valid responses, a multiple linear regression model was estimated with MHII as the dependent variable and fear barriers, nutrition awareness, athlete status, and stress frequency as predictors. The results indicate that fear barriers and stress frequency exert significant negative effects on mental health improvement, whereas nutrition awareness has a significant positive influence. Athlete status was not statistically significant. These findings demonstrate that psychological recovery associated with aerial sports participation can be modelled quantitatively. Based on these results, environmental design principles are derived to reduce fear barriers, regulate stress exposure, and reinforce nutritional awareness within rehabilitation settings.

Keywords: applied mathematics, sport psychology modelling, mental health index, aerial sports

1. INTRODUCTION

Depression and other mental health disorders have become increasingly widespread worldwide. These conditions vary according to age, gender, and personal circumstances, and rising prevalence rates highlight the need for effective preventive and rehabilitative strategies.

Mental health disorders are influenced by economic, social, and physical factors. Some cases are genetic, while others stem from chronic stress, loneliness, or social isolation (Dixit, 2023; Mills, 2011). Physical activity and lifestyle modification, particularly balanced nutrition, have been identified as important contributors to psychological well-being.

Aerial sports represent a distinctive category of physical activity that combines physical exertion, controlled risk exposure, and emotional regulation (Yao & Niu, 2023). Activities such as paragliding and climbing may contribute to stress reduction, enhanced self-confidence, and emotional resilience. However, participation may be limited by psychological barriers, particularly fear (Zighy, 2025).

While the psychological and physiological benefits of sport are widely discussed, fewer studies operationalize these relationships within a quantitative modeling framework (Hagger, & Hamilton, 2025). This study addresses that gap by constructing a measurable index of perceived mental health improvement and estimating the relative contribution of sport participation and lifestyle factors using regression analysis.

The study therefore seeks to:

- Model mental health improvement as a quantitative outcome.
- Estimate the influence of fear barriers, stress frequency, nutrition awareness, and athlete status.
- Derive environmental design implications from statistically significant predictors.

2. LITERATURE REVIEW

2.1 Sport and Mental Health

Dixit (2023) stated that depression may reduce engagement in sports, leading to declining mental health due to reduced encouragement, loneliness, and stress. Conversely, participation in sports can reduce stress, enhance self-esteem, and promote social support. Activities such as paragliding may assist individuals in confronting fears and overcoming isolation (Vella et al., 2025). These findings suggest a strong relationship between sport participation and psychological recovery. However, much of the literature remains descriptive rather than quantitatively modeled.

2.2 Nutrition and Psychological Outcomes

Mills (2011) identified the phenomenon of "hidden hunger," in which individuals consume sufficient calories but lack essential nutrients. Poor nutrition is associated with weakened immunity and increased risk of chronic disease. Social and economic factors influence dietary quality, reinforcing the need for nutritional awareness (Gombo er al., 2025).

Together, these perspectives indicate that both sport participation and nutrition awareness contribute to mental well-being. However, integrated quantitative modeling of these variables remains limited.

3. METHODOLOGY

3.1 Research Design

A mixed-method research design was adopted, combining quantitative and qualitative approaches. Quantitative data were collected through a structured survey, while qualitative insights were obtained through semi-structured interviews. This approach enabled both statistical modeling of mental health improvement and contextual understanding of environmental and psychological preferences.

The quantitative component formed the primary analytical foundation of the study, while qualitative findings were used to support interpretation and inform design implications.

3.2 Participants and Data Collection

A structured questionnaire consisting of 34 closed-ended questions was developed based on the study's objectives and conceptual framework. The questions were designed to measure demographic characteristics, psychological experiences, lifestyle factors, perceptions of aerial sports, and attitudes toward rehabilitation environments.

The survey was distributed electronically via WhatsApp and Telegram to individuals relevant to the research topic. Data collection took place over one week, from March 5 to March 11, 2025. A total of 108 valid responses were obtained.

Participants included individuals who had experienced psychological stress, individuals interested in improving lifestyle habits, athletes, adolescents seeking mental and nutritional awareness, and retirees who may be affected by lifestyle-related psychological stress. The inclusion of diverse participant groups ensured variation in stress levels, sport engagement, and nutritional awareness.

In addition to the survey, in-depth interviews were conducted with individuals who had experienced psychological stress. These interviews were unstructured, allowing participants to express their experiences and preferences freely. The flexible format enabled exploration of emerging themes related to spatial comfort, fear perception, environmental stress reduction, and lifestyle improvement.

3.3 Operationalization of Variables

To align the study with applied mathematical modeling, psychological and behavioral constructs were defined as measurable variables.

The following predictors were constructed:

- **Fear Barriers (F):** ordinal variable measuring hesitation toward aerial sports.
- **Stress Frequency (SF):** ordinal variable measuring frequency of psychological stress.
- **Nutrition Awareness (N):** ordinal variable reflecting awareness of the relationship between nutrition and mental health.
- **Athlete Status (AS):** dummy variable (1 = athlete, 0 = non-athlete).

Additional variables related to relaxation and aerial sports engagement were used in constructing the dependent variable.

Survey responses were numerically encoded using ordinal scales (e.g., Yes = 3, Maybe = 2, No = 1). Frequency responses were coded according to increasing intensity levels. This transformation enabled qualitative perceptions to be treated as quantitative variables suitable for statistical modeling.

3.4 Construction of the Mental Health Improvement Index (MHII)

A composite indicator termed the **Mental Health Improvement Index (MHII)** was constructed to represent perceived mental health improvement numerically.

The MHII was calculated as the average of selected normalized variables representing:

- Reduction of psychological stress
- Perceived emotional recovery
- Engagement or interest in aerial sports

This composite formulation allows mental health improvement to be expressed as a continuous numerical outcome rather than a descriptive concept. Higher MHII values indicate greater perceived improvement in psychological well-being.

By aggregating multiple dimensions into a single index, the MHII captures the multidimensional nature of mental health recovery while remaining analytically tractable for regression modeling.

3.5 Statistical Modeling Framework

To examine the influence of lifestyle and psychological factors on perceived mental health improvement, the following multiple linear regression model was specified:

$$MHII_i = \beta_0 + \beta_1 F_i + \beta_2 N_i + \beta_3 AS_i + \beta_4 SF_i + \varepsilon_i$$

Where:

- F_i= Fear Barriers
- N_i= Nutrition Awareness
- AS_i= Athlete Status
- SF_i= Stress Frequency
- ε_i= stochastic error term

Under the classical linear regression assumptions:

$$E(\varepsilon_i \mid F_i, N_i, AS_i, SF_i) = 0$$

$$Var(\varepsilon_i \mid F_i, N_i, AS_i, SF_i) = \sigma^2 < \infty$$
$$Cov(\varepsilon_i, \varepsilon_j) = 0 \text{for } i \neq j$$

the Ordinary Least Squares (OLS) estimator is obtained by minimizing the sum of squared residuals:

$$\min_{\beta_0,\beta_1,\beta_2,\beta_3,\beta_4} \sum_{i=1}^{n} (MHII_i - \beta_0 - \beta_1 F_i - \beta_2 N_i - \beta_3 AS_i - \beta_4 SF_i)^2$$

Under these assumptions, the OLS estimator provides the Best Linear Unbiased Estimator (BLUE) of the regression coefficients.

3.6 Qualitative Analysis

Qualitative data from interviews were analyzed using thematic analysis. This method allows identification of recurring patterns and themes without strict adherence to a single theoretical framework.

Thematic analysis focused on environmental preferences, fear perception, spatial comfort, sensory experiences, and integration of nature. These themes were later interpreted in relation to statistically significant predictors identified in the regression model.

The qualitative component therefore served a complementary role, supporting interpretation of quantitative findings and informing environmental design implications.

4. RESULTS

4.1 Survey Overview

A total of 108 valid responses were analyzed. The survey was designed to assess demographic characteristics, psychological stress frequency, lifestyle attributes, perceptions of aerial sports, and attitudes toward rehabilitation environments. Responses were numerically encoded using ordinal and categorical scales to enable quantitative modeling.

Demographic analysis indicated variation in age, occupation, and lifestyle characteristics. Participants reported experiencing psychological stress at varying frequencies, with work- and study-related pressures being common stressors. Most respondents expressed belief in the positive relationship between nutrition and mental health. Interest in aerial sports was present among many participants; however, fear of heights and perceived difficulty of access were frequently identified barriers.

Relaxation and psychological recovery were reported as primary motivations for interest in rehabilitation environments. Preferences for open spaces, natural settings, and calm design styles were consistently observed.

4.2 Quantitative Statistical Results

Descriptive Statistics

The respondents consisted of 28% female participants and 72% male participants, ensuring representation from both genders within the sample. Pie charts illustrating the distribution of individuals by Age and Occupation are presented in Figure 1, and descriptive statistics for the quantitative variables are provided in Table 1. The Mental Health Improvement Index (MHII) was computed as a composite indicator based on stress reduction, perceived relaxation, and aerial sports engagement.

Fear Barriers, Nutrition Awareness, and Stress Frequency were treated as ordinal variables, while Athlete Status was modeled as a dummy variable.

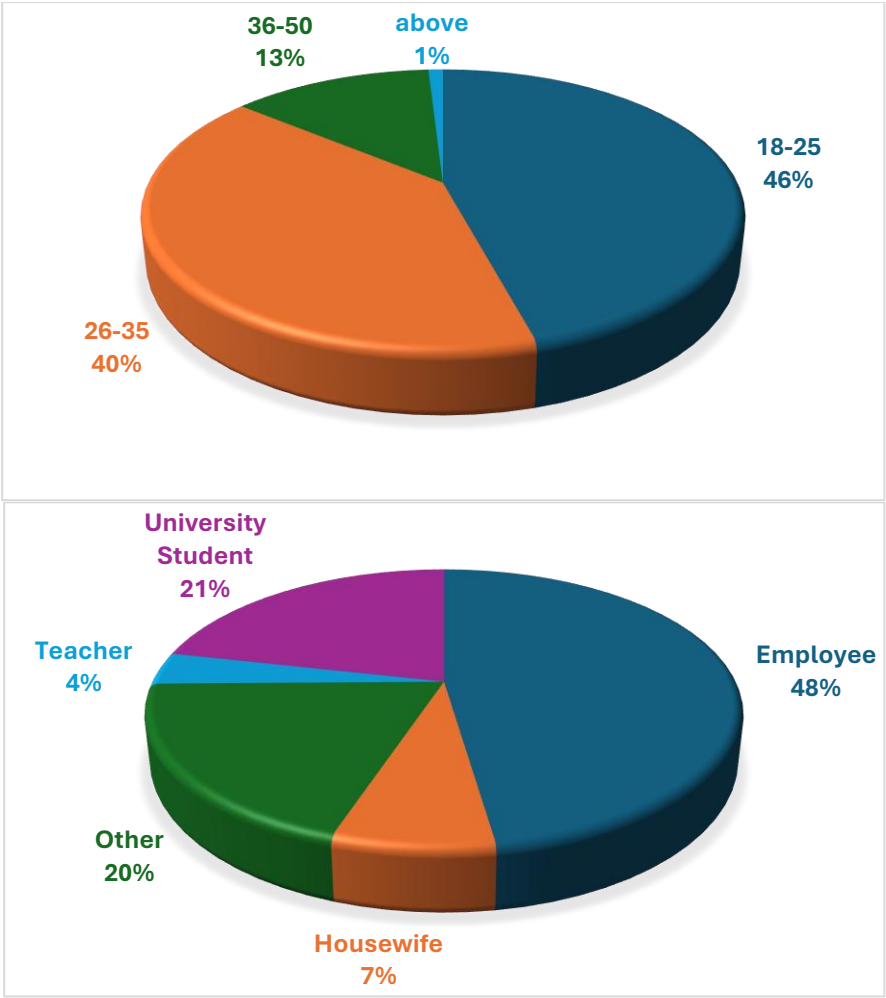


Figure1: Pie chart of participants by Age and Occupation

Table1: Descriptive Statistics

	N	Mean	Variance	Skewness	Kortosis
MHII	108	0.9998	0.3880	-0.102	-0.374
Fear_Barrier	108	2.4283	0.404	-0.801	-0.058
Nutrition_Awareness	108	3.0711	0.359	1.548	1.110
Stress_Frequency	108	2.8575	0.495	0.200	-0.128

Regression Analysis

To investigate the predictive relationship between lifestyle factors and perceived mental health improvement, a multiple linear regression model was estimated with MHII as the dependent variable:

$$MHII_i = \beta_0 + \beta_1 F_i + \beta_2 N_i + \beta_3 AS_i + \beta_4 SF_i + \varepsilon_i$$

The regression model was statistically significant, as indicated by the ANOVA results in Table 2.

Table2: ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	13.039	4	3.260	11.794	0.000
Residual	28.468	103	0.276		

Total	41.507	107			

The F-statistic (11.794, $p < 0.001$) confirms that the predictors jointly explain variation in MHII.

The estimated coefficients are presented in Table 3.

Table3: Coefficients

Model	Coefficients	Std. Error	t	sig
Intercept	1.245	0.312	3.99	0.000
Fear Barrier (F)	-0.284	0.072	-3.94	0.001
Nutrition Awareness (N)	0.356	0.081	4.39	0.003
Athlete Status (AS)	0.198	0.067	1.62	0.109
Stress Frequency (SF)	-0.241	0.069	-3.49	0.001

The estimated coefficients (Table 3) reveal:

- **Fear Barriers (β = -0.284, p = 0.001)** significantly reduce MHII.
- **Stress Frequency (β = -0.241, p = 0.001)** significantly reduces MHII.
- **Nutrition Awareness (β = 0.356, p = 0.003)** significantly increases MHII.
- **Athlete Status (β = 0.198, p = 0.109)** is not statistically significant.

The fitted regression equation is:

$$\widehat{MHII} = 1.245 - 0.284F + 0.356N - 0.241SF$$

Fear Barriers and Stress Frequency have significant negative effects on mental health improvement, while Nutrition Awareness has a significant positive effect. Athlete Status is not statistically significant.

These findings suggest that mental health improvement is more strongly influenced by psychological barriers and lifestyle awareness than by athletic identity alone.

Diagnostic testing was conducted to verify regression assumptions. The Shapiro–Wilk test indicated no significant deviation from normality of residuals ($p > 0.05$). Variance Inflation Factors (VIF < 5) suggested no multicollinearity among predictors. The Durbin–Watson statistic (DW = 1.70) indicated no serious autocorrelation.

These results support the robustness and validity of the regression estimates.

4.3 Qualitative Interview Findings

In-depth interviews were conducted with individuals who had experienced psychological stress. Thematic analysis identified recurring themes related to environmental comfort, emotional regulation, and spatial preferences.

Participants emphasized the importance of calm and neutral color schemes, along with natural lighting and ventilation, as key elements contributing to psychological comfort. They also highlighted the value of clearly separating quiet indoor spaces from more active outdoor areas, allowing individuals to transition gradually between relaxation and engagement. The integration of nature was consistently identified as essential for enhancing emotional well-being, while gradual exposure to physically demanding activities was viewed as important for reducing fear and supporting confidence development.

Fear and hesitation toward aerial sports were discussed in relation to perceived safety and lack of familiarity. This qualitative insight aligns with the statistically significant negative effect of Fear Barriers identified in the regression model.

Participants also expressed strong belief in the connection between healthy nutrition and mental well-being, reinforcing the positive coefficient of Nutrition Awareness in the quantitative analysis.

Overall, qualitative findings supported the quantitative results by highlighting stress regulation, gradual exposure, and lifestyle awareness as central components of perceived psychological recovery.

5. DISCUSSION

5.1 Interpretation of Quantitative Findings

The regression results demonstrate that mental health improvement, as measured by the Mental Health Improvement Index (MHII), is significantly influenced by psychological and lifestyle variables associated with aerial sports and nutrition. Fear Barriers and Stress Frequency exert statistically significant negative effects on MHII, while Nutrition Awareness shows a statistically significant positive contribution. Athlete Status was not statistically significant.

These findings suggest that perceived psychological recovery is less dependent on athletic identity and more strongly influenced by the reduction of psychological fear and ongoing stress. Individuals who experience frequent stress or high levels of hesitation toward aerial sports report lower perceived mental health improvement. Conversely, greater awareness of the relationship between nutrition and mental health contributes positively to psychological well-being.

The results confirm that mental health improvement is not a binary outcome but a multidimensional process shaped by interacting psychological and behavioral factors. This supports the need for integrated interventions that simultaneously address stress regulation, fear reduction, and lifestyle awareness.

5.2 Environmental and Design Implications Derived from the Model

The quantitative findings have direct implications for the design of rehabilitation environments. Since Fear Barriers significantly reduce MHII, spatial environments should be structured to gradually reduce perceived risk. Rather than immediately exposing individuals to high-intensity aerial sports, a phased progression of activities may be implemented. This aligns with participants' expressed need for safety, reassurance, and controlled engagement.

Stress Frequency also negatively affects mental health improvement, indicating that environmental conditions must support stress reduction. Participants emphasized the importance of natural landscapes, open spaces, and calm environments. Quiet zones for relaxation, sensory areas incorporating water and natural scents, and spaces designed for contemplation may help regulate stress levels and enhance recovery.

Nutrition Awareness demonstrated a positive effect on MHII. This suggests that environments promoting healthy nutrition, such as educational food areas, farms, or accessible healthy dining spaces,may reinforce psychological improvement. Participants strongly believed in the connection between nutrition and mental health, highlighting the importance of raising awareness and providing healthy food options.

The qualitative findings reinforce these implications. Interviewees consistently preferred natural and calm design styles, separation between quiet and active areas, and integration of nature. These preferences align with the statistical evidence that stress regulation and gradual fear exposure are central to mental health improvement.

Thus, environmental design should not be viewed as purely aesthetic, but as a structured response to measurable psychological variables.

5.3 Conceptual Design Translation: The Cochlea as a Spatial Metaphor

The quantitative analysis indicates that mental health improvement is progressive and cumulative. Reductions in fear and stress correspond to incremental increases in perceived psychological recovery. Translating this layered progression into spatial design requires a conceptual framework representing balance, adaptation, and transformation.

The cochlea, a spiral-shaped structure within the inner ear responsible for balance and sound processing, provides a symbolic representation of this process. The spiral form reflects movement from internal states to external engagement, mirroring the journey from psychological stress toward equilibrium.

In the proposed conceptual translation, the spiraling pathways of the cochlea represent the progression of spaces within the rehabilitation environment. Movement begins in internal, calm zones dedicated to relaxation and stress reduction, and gradually transitions toward more active zones, including aerial sports engagement areas. The curvature applied in architectural form symbolizes adaptation and balance, reinforcing the idea that recovery is a gradual and structured process.

The cochlea metaphor therefore aligns with the modeled trajectory of mental health improvement captured by the MHII. Just as the cochlea adapts to sound and maintains equilibrium, the rehabilitation environment adapts to varying levels of fear and stress, guiding individuals toward improved psychological stability.

5.4 Importance and Future Research Directions

This study demonstrates that psychological recovery associated with aerial sports participation can be modeled quantitatively. The construction of the MHII provides a measurable framework for evaluating the influence of fear barriers, stress frequency, and nutrition awareness on perceived mental health improvement.

The integration of applied mathematical modeling with environmental design contributes to the field of sport psychology and rehabilitation by translating statistical findings into practical spatial strategies. Rather than designing environments based solely on intuition or aesthetic preference, this approach derives design principles from empirically identified predictors.

Future research may expand this modelling framework by incorporating longitudinal data to examine how the MHII evolves over time, thereby capturing dynamic changes in psychological recovery. Additional psychological variables, such as resilience, social support, and perceived safety, could be included to enhance the explanatory power of the model. Comparative analyses across different categories of sport participation beyond aerial sports would allow evaluation of whether similar patterns hold in other physical activity contexts. Further investigation of potential non-linear relationships or interaction effects between stress and fear variables may also provide deeper insight into the mechanisms underlying mental health improvement. Finally, validation of the MHII in larger and more diverse populations would strengthen its generalizability and establish its broader applicability within sport-based rehabilitation research.

Further development of quantitative indices may also support predictive modeling of rehabilitation outcomes and optimization of sport-based intervention programs.

By bridging quantitative modeling and environmental application, this study highlights the importance of measurable frameworks in designing effective sport-based rehabilitation environments. The findings underscore that mental health improvement is influenced by structured exposure, stress regulation, and lifestyle awareness, and that these factors can be systematically modeled and translated into design practice.

6. CONCLUSION

This study examined the design of a rehabilitation environment that integrates aerial sports and nutritional awareness through a quantitative modeling framework. In response to the increasing prevalence of mental health challenges and the limited availability of facilities that combine psychological therapy, healthy nutrition, and physical activity, the study operationalized mental health improvement as a measurable outcome.

The construction of the Mental Health Improvement Index (MHII) enabled psychological recovery to be expressed numerically rather than descriptively. Using multiple linear regression, the analysis demonstrated that Fear Barriers and Stress Frequency significantly reduce perceived mental health improvement, while Nutrition Awareness contributes positively. Athlete Status was not statistically significant, indicating that mental health benefits are not restricted to individuals who already identify as athletes.

These findings highlight the strong connection between mental health, physical activity, and nutrition, and confirm that these factors collectively influence overall well-being. The results emphasize the importance of reducing fear barriers, regulating stress exposure, and promoting nutritional awareness within sport-based rehabilitation environments.

The study further translated these statistical findings into environmental design implications. Spaces that balance tranquility and activity, integrate nature, and provide gradual engagement in aerial sports may support psychological recovery. The conceptual design inspiration drawn from the cochlea symbolizes balance, adaptation, and progressive transformation, reflecting the cumulative nature of mental health improvement identified through the MHII.

This research contributes to applied sport mathematics by demonstrating how psychological constructs associated with aerial sports participation can be quantified and modeled. By linking statistical analysis to environmental application, the study underscores the importance of measurable frameworks in guiding rehabilitation strategies.

Ultimately, the goal is to support individuals in achieving improved emotional stability, enhanced self-confidence, and renewed engagement with life through structured sport participation and healthy lifestyle practices. The integration of quantitative modeling and environmental design offers a foundation for future research and development of comprehensive sport-based rehabilitation programs.

7. References

Dixit. Prince V., The impact of sports participation on the prevention and management of depression: a comprehensive study", 10 October 2023. Retrieved march 12 2025, from https://www.researchgate.net/publication/374440728_THE_IMPACT_OF_SPORTS_PARTICIPATION_ON_THE_PREVENTION_AND_MANAGEMENT_OF_DEPRESSION_A_COMPREHENSIVE_STUDY.

Gomboş, L., Milon, A. G., Nuţ, R. A., Dragomir, C., & Gherman, A. A. (2025). Sport, Culture, and Purchasing Power: Influences on Food Markets and Consumption. Amfiteatru Economic, 27(69), 367-383.

Hagger, M. S., & Hamilton, K. (2025). Testing theory predictions and intervention mechanisms of action in sport and exercise psychology using meta-analysis. Sport, Exercise, and Performance Psychology, 14(1), 96.

Mills, Dora A., Poor Nutrition Amidst Plenty, January 2011. Retrieved march 12 2025, from https://www.researchgate.net/publication/361585362_Poor_Nutrition_Amidst_Plenty.

Vella, S. A., Liddelow, C., Rice, S. M., Keegan, R., Hall, K., Jones, M. A., ... & Swann, C. (2025). Development of Australian mental health guidelines for community sport. British journal of sports medicine, 59(9), 659-666.

Yao, Y., & Niu, X. (2023). Construction of a physical fitness evaluation index system and model for high-level freestyle skiing aerials athletes in China. Plos one, 18(12), e0295622.

Zighy Bay Group, Living in harmony at Zighy Bay, 2025. Retrieved February 16th 2025, from https://www.sixsenses.com/ar/hotels-resorts/middle-east-africa/oman/zighy-bay/sustainability/

MEASURING EXCITEMENT IN THE CLOSING STAGES OF THE MASTERS GOLF

Graham Pollard [a,d] , Anthony Bedford [b] , Erica Mealy[b]
[a] *University of Canberra*
[b] *University of the Sunshine Coast*
[d] *Corresponding author: grahamhpollard@gmail.com*

Abstract

This paper develops a probabilistic framework for measuring excitement and surprise in the closing stages of the Masters Golf Tournament. Excitement is defined as the expected magnitude of the change in a player's probability of winning the tournament resulting from the next event, such as the completion of a hole. Using forward recurrence methods and realistic birdie, par and bogie probabilities for leading players, the model evaluates how winning probabilities evolve as the final holes are played. Illustrative examples show how excitement increases when multiple players remain in contention and how course characteristics influence this dynamic. A complementary measure of surprise is also introduced to quantify the retrospective impact of unusual outcomes. The approach allows excitement to be updated continuously during play and has potential applications for broadcast analytics and real-time commentary.

Keywords: Excitement, golf, recursive probabilities

1. INTRODUCTION

The Masters is generally regarded as the premier golf tournament in the world and is held over four days on a Thursday to Sunday in early April each year. It is a particularly exciting event if, near the very end of play on the Sunday, there are two or more players who each have a quite reasonable probability of winning the event. If, say with half an hour of play left on the Sunday, it is quite clear who is going to win, the event is not as exciting as it otherwise might be.

If two or more players are potential winners after almost four days of play, it can be argued that they are effectively equal, at least probabilistically. Thus, it would appear reasonable to assume that such players have the same probability of scoring a birdie (or a par, or a bogie) on say the 18th hole (or the 17th, etc). Given their relative overall scores at the closing stages on the Sunday (e.g. 10 under, 11 under, etc) and their likely scores on the final hole(s), it is possible to use forward recurrence methods to evaluate each player's probability of winning the event.

Note that it is possible to measure the excitement as it builds up or dissipates at the final stages of the event. Also, it is possible to compare the excitement of the final few holes from one year to previous or later years.

2. LITERATURE REVIEW

Pollard (2017) studied one of the most exciting tennis matches ever played: the five-set final of the 1980 Wimbledon Men's Final between McEnroe and Borg. McEnroe won the 4th set 18-16 in the tiebreak game to take the match to the 5th set. Pollard (2017) defined the excitement of a particular point within a set of tennis as the expected value of the absolute size of the change in a player's probability of winning the set as a result of that point being played. He showed how the mean and variance of the total excitement in a set of tennis can be evaluated, thus making it possible to compare the excitement of two or more completed sets of tennis. This definition of the excitement associated with a point in the tennis context can be applied directly to the excitement of a hole in the final few holes of the Masters. This is the approach taken in this study.

Jason Lusk (2022) noted the relative difficulty of the 18 holes at The Masters. Table 1indicates the difficulty of the final six holes in the Masters, and (more than) covers the holes that are relevant for this study. As displayed, holes 13 and 15 are the easiest on the course, whilst holes 14, 16, 17, 18 have 'average difficulty'. Note that the course has changed a little over time. For example, the 13th hole was lengthened by 35 yards in 2023.

Note that these averages over all players date back to 1934, that the averages for today's players have clearly improved. Also, note that averages for the final players on the Sunday would be even better than the averages of all players in the event.

Hole	Rank Historical 18 = easiest	Length	Par	Average Historical	Lowest Average
13	18, easiest	510	5	4.77	4.474 in 1991
14	8, 11th easiest	440	4	4.17	3.936 in 2011
15	17, 2nd easiest	550	5	4.77	4.505 in 1991
16	11, 8th easiest	170	3	3.14	2.875 in 2020
17	9, 10th easiest	440	4	4.16	3.949 in 1996
18	7, 12th easiest	465	4	4.23	4.014 in 2001

Table 1: Lusk data … Golfweek https://golfweek.usatoday.com/lists/2023-masters-augusta-national. (See Jason Lusk, April 5, 2022. Golf News Net. https://thegolfnewsnet.com/golfnewsnet and the Golf News Net dated 6 April 2023 and Golfweek https://golfweek.usatoday.com/lists/2023-masters-augusta-national.)

Hole	Rank	Length(Yds)	Par	Average
13	16	545	5	4.852
14	12	440	4	4.046
15	15	550	5	4.933
16	14	170	3	3.004
17	8	440	4	4.205
18	2	465	4	4.389

Table 2: Difficulty of holes

Focusing on the final holes and the leading players on the Sunday, the birdie/par/bogie distribution for these (virtually equal) players might be as shown in Table 3.

Hole/probability	Par 5	Par 4	Par 3
Birdie	0.5	0.2	0.2
Par	0.45	0.7	0.6
Bogie	0.05	0.1	0.2

Table 3: Birdie, Par and Bogie probabilities for the players on the final holes.

RECENT HISTORY OF PLAYOFFS
Since 2000 there have been playoffs on 7 occasions.
2003…Mike Weir beat Len Mattiace…on the first extra hole.
2005…Tiger Woods defeated Chris DiMarco on first playoff hole.
2009…Three players…Angel Cabrera won a playoff, beating Kenny Perry (after a second extra hole) and Chad Campbell (after first extra hole).
2012…Bubba Watson beat Louis Oosthuizen (on the second extra hole, the 10th).
2013…Adam Scott beat Angel Cabrera on the second playoff hole.
2017...Sergio Garcia beat Justin Rose on the first extra hole.
2025…Rory McIlroy beat Justin Rose on the first playoff hole.

In this paper the word ***excitement*** is used when a player has a typical result on a hole such as a par, a birdie or a bogie, whilst the word ***surprise*** is used when a player has a more unusual result such as an eagle or a double bogie. However, the measurement of ***excitement*** and the measurement of ***surprise*** is calculated using the same mathematical concept. Note that eagles and double bogies, being ***surprise*** outcomes, are handled mathematically as they occur.

EXAMPLE 1
Suppose we have two equal players A and B who are both at 10 under par on the 17th tee. Suppose they are the only players who can possibly win the Masters (and that the leader in the clubhouse has no real chance of winning). Assume *for simplicity* in this example that both players A and B have values P(birdie) = P(par) = P(bogie) = 1/3 on both holes 17 and 18.
It is clear that with both (equal) players about to tee off on the 17th tee,

$$P(\text{A wins the Masters}) = P(\text{B wins the Masters}) = 0.5. \quad (0)$$

Now suppose the next critical thing that happens is player A getting a birdie on the 17th hole, going to 11 under on the 18th tee. Then, if player B gets a bogie on the 17th hole with probability 1/3, he goes to 9 under on the 18th hole, and player A must win the Masters. If player B gets a par on the 17th with probability 1/3, he goes to 10 under on the 18th tee and player A wins with probability (2/3 + (1/3*0.5) = 5/6. Also, if player B gets a birdie on the 17th hole player with probability 1/3, he goes to 11 under on the 18th tee, and player A wins with probability 0.5. Hence, given player A gets a birdie on the 17th hole,

$$P(\text{A wins overall}) = 1/3*1 + (1/3)*(5/6) + (1/3)*(1/2) = 7/9 \quad (1)$$

The excitement of player A getting a birdie on the 17th is thus

$$\text{Excitement(birdieA)}=(1/3)*(1\text{-}0.5) + (1/3)*(5/6\text{-}0.5) + (1/3)*(1/2\text{-}1/2) = 5/18 \quad (2)$$

EXAMPLE 2 Excitement on the 18th hole

Consider an example of the excitement with the final three (equal and leading) players (A, B and C) on the 18th tee, all on 10 under, and with the leader in the clubhouse out of contention. To keep the calculations simple, it is assumed that for each of these players, P(Par) = 0.7 and P(Birdie) = 0.3. (i.e. Each of them cannot get a bogie.) [Note that whilst there is only ever in practice (at most) two players on a tee at the Masters, the situation with 3 players on a tee can be evaluated mathematically.]
Then there is a balanced likelihood of A or B or C only winning at the 18th, as such

$$P(i \text{ only wins Masters at } 18^{th}) = P(\text{Birdie})*P(\text{Par})*P(\text{Par}) = 0.3*0.7*0.7 = 0.147 \quad (3)$$

where $i \in (A, B, C)$ exclusively. Generalisable, if we define $P(Par|player = i) = \pi_i$ and $P(Birdie|player = i) = \beta_i$ then we can have

$$P(i \text{ only wins Masters at 18th}) = \pi_i\beta_j\beta_k \quad (3b)$$

Similarly, if only two players proceed to a playoff, then the third player makes par (vs. Birdies), calculated

$$P(j \text{ only win the } 18^{th} \text{ and go to playoff hole}) = 0.3*0.3*0.7 = 0.063 \quad (4)$$

where $j \in \{(A \cap B), (B \cap C), (A \cap C)\}$

Finally,

$$P(\text{A, B and C all win on the } 18^{th} \text{ and go to playoff}) = 0.7*0.7*0.7 + 0.3*0.3*0.3 = 0.37 \quad (5)$$

$$\text{Excitement of A on } 18^{th} = 0.147*\text{Mod}(1\text{-}1/3) + 0.147*\text{Mod}(1/3\text{-}0) + 0.147*\text{Mod}(1/3\text{-}0) + 0.063*\text{Mod}(0.5\text{-}1/3) + 0.063*\text{Mod}(0.5\text{-}1/3) + 0.063*\text{Mod}(0\text{-}1/3) + 0.37*(1/3\text{-}1/3) = 0.238. \quad (6)$$

As the players are equal, Excitement B on 18th = Excitement C on 18th= Excitement A on 18th, and so Total Excitement on 18th= 0.714.

More Generally, the above example was repeated with the same parameters for 2, 4 and 5 players on the 18th tee, even though in practice there is only ever 2 players on that tee on the final day at the Masters. (Note that in some other tournaments there might be 3 or 4 players on each tee.) The results are given in Table 4.

Number of equal players	Total Excitement	Excitement per player
2	0.42	0.21
3	0.714	0.238
4	0.9198	0.22995
5	1.06386	0.212772

Table 4: The total excitement with a number of equal players on the 18th tee.

Note that the excitement per player is maximum when there are 3 players on the 18th tee, as may have been anticipated. A generalisable result could be determined off historical results for the players, utilising the following notation

EXAMPLE 3
The excitement with two players at the 19th (Playoff) hole
Consider a playoff between two players on the 19th hole. The order of the playoff holes is the 18th, 10th, 18th, 10th, 18th, until a winner is determined. The 18th hole 'Holly' is a 465 yd par 4 whilst the 10th hole 'Camellia' is a 495 yd par 4 and is *historically* the hardest hole on the course. The present data suggest that these two holes have reasonably similar difficulty.
For each of these two holes and these leading players, it is assumed P(Birdie) = 0.2, P(Par) = 0.7 and P(Bogie) = 0.1. These are representative probabilities for such leading players. It is assumed that the outcome for player A (birdie, par or bogie) is independent of the outcome for player B.

Probability		B birdie	B par	B bogie
	Product	0.2	0.7	0.1
A birdie	0.2	0.04	0.14	0.02
A par	0.7	0.14	0.49	0.07
A bogie	0.1	0.02	0.07	0.01

Table 5: Analysis of a playoff on the first playoff hole (19th) with two players A and B.

$$P(\text{A wins on the } 19^{th}) = P(\text{B wins on the } 19^{th}) = 0.14+0.02+0.07 = 0.23 \quad (7)$$

$$\text{Excitement A} = 0.23*\text{Mod}(1-0.5) + 0 + 0.23*\text{Mod}(0-0.5) = 0.23 \quad (8)$$

Correspondingly, Excitement B = 0.23.
Total Excitement at 19th hole = Excitement A + Excitement B = 0.46.
Since the other playoff hole (the 16th) is assumed to have the same player characteristics,
Total Expected Excitement until the end of play = $0.46 + 0.54*0.46 + (0.54^2)*0.46+\ldots = 1$.

EXAMPLE 4 - The excitement with three equal players on the 19th (Playoff) hole
A natural extension is to consider three equal players A, B and C with parameters as previously states of (0.2, 0.7, 0.1) playing off on the 19th hole (as in example 3). Using the same approach, Playoff on the 1st playoff hole (physically the 18th) with 3 equal players A, B and C is thus

P(A wins outright on the 19th hole) = 0.135.

P(A and B, or A and C go to 20th hole to continue the playoff (2 players only)) = 0.162.
P(A, B and C all go through to 20th hole = 0.352.
P(B or C wins outright on 19th OR B and C only go on to 20thhole) = 0.351.
These 4 probabilities above add to 1.0.
Excitement A = (0.135*(1-1/3)) + (0.162*(1/2-1/3)) + (0.351*(1/3-1/3)) + (0.351*(1/3-0)) = 0.234.
Excitement A = Excitement B = Excitement C, as the players are equal (i.e.. identical).
Total Excitement on this 19th hole= 0.702 (which is a little bigger than 0.46*1.5).
Note that the total expected excitement until the end of play could be determined. This is a little complicated.
Note that on the 20th hole we have either 3 players in contention (a repeat of what we have just considered), or just two players in contention (which is a repeat of the case considered earlier). Thus, it is possible to calculate the total expected excitement until the end of play, starting with three players on the 19th tee.

EXAMPLE 5 – Four players on same score but differing holes
Suppose we have 4 equal players A, B, C, and D in contention of winning. Suppose they are on the 15th, 16th, 17th and 18th tees respectively, and are all 10 under, and the leader L (also an equal player) is in the clubhouse is also 10 under. Suppose all (equal) players have a birdie, par, and bogie probability of 0.2, 0.7 and 0.1 respectively (note that birdie more likely than bogie). Then, it can be shown using forward recurrence methods that

$$P(\text{A wins}) > P(\text{B wins}) > \text{Prob}(\text{C wins}) > \text{Prob}(\text{D wins}) > P(\text{L wins}) \quad (*)$$

This is essentially because A has more holes to play than B, and B has more than C,…etc,
and a birdie is more likely than a bogie on each hole (such there is a dominance as long as P(birdie)>P(bogie).

EXAMPLE 6 – Consideration of three players from example 5
We assume the birdie, par and bogie probabilities on each of the 17th and 18th holes for each player are 0.2, 0.7, 0.1. (i.e. they are equal players and the holes are equally difficult). We consider the situation where the leader L, is in the clubhouse at 10 under, player B is on the 18th tee at 10 under, and player A is on the 17th tee also at 10 under. As noted in (*) the likelihood of winning is greatest for the player with more resources remaining. The forward recurrence method has been used to determine the distribution of the players' scores after 18 holes have been completed. This is given in Table 6.

Prob After 18 holes	A score	B score	L score	A only wins	B only wins	L only wins	A and B only win	A and L only win	B and L only win	A, B and L only win
12 under	0.04			0.04						
11 under	0.28	0.2		0.224	0.136		0.056			
10 under	0.53	0.7	1			0.015		0.053	0.105	0.371
9 under	0.14	0.1								
8 under	0.01									

Table 6: Distribution of players' scores after 18 holes using forward recurrence.

Note that the numbers in the last seven columns of Table 6 cover all circumstances and therefor sum to 1.0. As the Leader is in the clubhouse au 10 under, A and B cannot win at 9 under or less although there is a slight a reasonable chance that one or both of these instances can occur. Probabilities are omitted in Table 6 for impossible events.
After a playoff of the relevant 2 or 3 people, we have
P(A Wins) = P(A only wins|12u)+P(A only wins|11u) + P(A&Bonlywin|11u)/2 + P(A&Lonlywin|10u)/2 + (P(A&B&L only win |10u)/3

P(A Wins) = 0.04+0.224+0.056/2+0.053/2+0.371/3 =0.4421666

P(B wins) = P(B only wins|11u) + P(A&B only wins|10u)/2 + P(B&Lwin|10u)/2 + P(A&B&Cwin|10u)/3
=0.3401666,
and similarly P(L wins) = 0.2176666, and these three values add to unity.

If we repeat the above recurrence analysis for the case when A is 11 under on the 18th tee and B is 10 under on the 18th tee, we have P(A wins) = 0.858333, P(B wins) = 0.113333, P(L wins) = 0.028333.

If we repeat the above recurrence analysis for the case when A is 10 under on the 18th tee and B is also 10 under on the 18th tee, we have P(A wins) = 0.378333, P(B wins) = 0.378333, P(L wins) = 0.243333.

If we repeat the above analysis for the case when A is 9 under on the 18th tee and B is 10 under on the 18th tee, we have P(A wins) = 0.056666, P(B wins) = 0.526666, P(L wins) = 0.416666.

It follows that
Excitement(A) = 0.2*(0.858333-0.442166)+0.7*(0.442166-0.378333)+0.1*(0.442166-0.056666)
= 0.1664665
Excitement (B) = 0.2*(0.340166-0.113333)+0.7*(0.378333-0.340166)+0.1*(0.526666-0.340166)
=0.090733
Excitement (L) = 0.2*(0.217666-0.028333)+0.7*(0.243333-0.217666)+0.1*(0.41666-0.217666)
=0.0757335 and so
And Total Excitement = 0.332933.

EXAMPLE 7

Suppose we do the same as in Example 6 above, but let's imagine that Hole 17 is an 'easier' hole (like the 15th hole), where the birdie, par and bogie probabilities are 0.4, 0.5 and 0.1 respectively. Forward recurrence methods lead to Table 7.

Score	A	B	L	A only	B only	L only	A and B only	A and L only	B and L only	A,B and L
12 under	0.08	0		0.08						
11 under	0.38	0.2		0.304	0.108		0.076			
10 under	0.41	0.7	1			0.013		0.041	0.091	0.287
9 under	0.12	0.1								
8 under	0.01									

Table 7: Easier 17th hole.

After a playoff of the relevant 2 or 3 people, we have
P(A Wins) = 0.08+0.304+0.076/2+0.041/2+0.287/3 =0.5381666 ; P(B wins) = 0.2871666; and P(L wins) = 0.1746666, and these add to unity.
It follows that
Excitement(A) = 0.4*(0.858333-0.5381666)+0.5*(0.538166-0.378333)+0.1(0.538166-0.056666)
= 0.2561334
Excitement (B) = 0.4*(0.2871666-0.113333)+0.5*(0.378333-0.2871666)+0.1*(0.526666-0.287166)
= 0.1390665
Excitement (L) = 0.0585332+0.034333+0.0242 = 0.117066, and so
Total Excitement = 0.512266
Thus, it can be seen that with the increased probability of a birdie on the 17th from 0.2 to 0.4, the total excitement has increased from 0.332933 to 0.512266.

EXAMPLE 8

Measuring 'Surprise' and Subjective probabilities

In the analyses above it has been assumed that a birdie, par and bogie are the only possible outcomes on a hole for a professional player in the Masters. This is a reasonably basic assumption, however occasionally a player has an '***surprise'*** eagle (-2) or a double boggie (+2). We show by way of an example how this can be handled *when it does in fact occur*.

Suppose we have the two leading players on the (par 4) 18th tee on the same score. Now suppose a short while later both players have played two shots. Suppose player A has done a marvellous second shot and is about 7 feet from the hole with a probability of (say) 0.8 of sinking the putt for a birdie, whilst player B is on the edge of the green with very little chance (say zero) of a birdie, a high likelihood (say 0.9) of a par, and a small chance (0.1) of a bogie. The following table shows the joint probability for the players' scores.

	Joint Prob		Player B Birdie	Par	Bogie
				0.9	0.1
	Birdie	0.8		0.72	0.08
Player A	Par	0.2		0.18	0.02
	Bogie				

Table 8: An example of a 'surprise' outcome

Note that prior to the players A and B commencing play the 18th hole, Prob(A wins) = Prob(B wins) = 0.5.
Now, after each player has had two strokes on this (par 4) 18th hole, the probability Player A wins the Masters (after a possible playoff) is equal to (0.72+0.08+0.02) + 0.18*0.5 = 0.91.
Thus, Surprise(A) = 0.91-0.5 = 0.41, and Surprise (B) = Mod(0.09-0.5) = 0.41, giving a Total Surprise of 0.82.
This example demonstrates that it can be possible to update an excitement measure whilst a hole is being played. It is called a 'surprise' measure and occurs when a hole is partially completed and something

somewhat unusual has happened. Note that this application of 'surprise' makes use of *subjective probabilities* for each player. The 'excitement' measure of 0.46 for this hole increased to a 'surprise' measure of 0.82 after both players had played just 2 strokes each.
The capacity for a skilled commentator to measure the 'surprise' level whilst a hole is partially completed is a useful aspect of this work. Further, it could be useful to think of excitement as a 'forward-looking' measure, and surprise as a 'backward-looking' measure given some relatively unusual event that has just occurred.

EXAMPLE 9 A second example of Measuring 'surprise'.
Here we consider a slightly modified example of the above. Again, suppose both players are even on the 18th tee. Now suppose the players have played two strokes each and are both near the edge of the 18th green very close to each other and about the same distance from the pin. They would appear to have an equal chance of winning the Masters, as given in the following Table.

	Probability
P(A wins Masters)	0.5
P(B wins Masters)	0.5

Table 9: Balanced likelihoods

Suppose player A plays first as he is very slightly further away from the hole than player B. Now suppose player A chips into the hole. Player B is most unlikely (say probability 0.02) to do the same, and so we have (say).

	Probability
P(A wins Masters on 18th green)	0.98
P(Players A and B go to 19th hole)	0.02

Table 10: Surprise! I chipped to nearly in

The measure of surprise is (0.99-0.5) for player A and (0.5-0.01) for player B, totalling 0.98.
Note that this is a probability measure of change *after* the event has occurred…a 'backwards-looking' measure.

EXAMPLE 10
In the next few examples we show that it is possible for the total excitement to be 1 or even a little greater. Consider two players A and B on the 19th playoff tee. Suppose player A is a very consistent player and only ever gets par (never a birdie or a bogie) and suppose player B only ever gets a birdie or a bogie, each with probability 0.5. Then, clearly the winner is determined on the 19th green, and
P(A wins) = P(A gets par and B gets bogie) = 1*0.5 = 0.5, and correspondingly P(B wins) = 0.5.
Excitement A when player A is about to play the 19th =
P(B gets bogie)*Mod(1-0.5) + P(B gets birdie)*Mod(0-0.5) = 0.5.
Correspondingly, the excitement when player B is about to play the 19th =
P(B gets bogie)*Mod(0-0.5) + P(B gets birdie)*Mod(1-0.5) = 0.5,
And so the total excitement is equal to 1.

EXAMPLE 11
Suppose we modify Example 10 and assume there are two players B1 and B2 who each get only a birdie or a bogie, each with probability 0.5, whilst, as above, Player A always gets par. Then it can be shown that Excitement B1 = Excitement B2 = 12/32, and Excitement A = 12/32, so that the Total Excitement = 36/32 = 1.125.
EXAMPLE 12
If we assume there are now 3 players B1, B2 and B3 with the same characteristics as in the above two examples, it can be shown that P(A wins) = 1/8, P(B1 wins) = P(B2 wins) = P(B3 wins) = 14/48, and Excitement A = 14/64, Excitement Bi = 14/48, I = 1, 2 and 3, and so Total Excitement = 105/96. Note that this total Excitement is a little less than in the previous example.

4. CONCLUSION

The excitement can be updated through hardcoding as play continues – as an online platform (even excel if punished enough) could do such calculations. Players can be included as they come into contention or deleted

as they fall from contention. Thus, the total (momentary) excitement can be measured hole by hole and noted as it ebbs and flows – certainly of value to those numerically interested. As strong extension could be the simulation of results to the densities generated here. The backward recursion lends itself to precise multi-dimensional in play contingency strategy and simulated sampling from player based prior distributions which lends us to more complex modelling as seen in Shepherd et. al. (2010).

Given these measures of excitement and surprise are based on probabilities, they are smallish values but might be multiplied by (say) 100 for broadcast viewers to make them more 'user friendly' (akin to percentages.) Such an approach could be implemented in a very straight forward fashion, and has obvious possibilities as an extension to bookmakers and punters alike for in-play trading.

References

Shepherd, R., Bedford, A., Schembri, A. A two-stage simulation to predict medalists in pistol shooting. Australasian Conference on Mathematics and Computers in Sport (10 M&CS), 10th (Darwin, Australia, 05-Jul-2010–07-Jul-2010) MathSport (ANZIAM)

Pollard G. H., 2017, Measuring Excitement in Sport. Journal of Sports Analytics.
Lusk, J., 2022.

Vecer J., Ichiba T. and Laudanovic M. 2007. On probabilistic Excitement in Sports Games. Journal of Quantitative Analysis in Sports, 3(3), article 6.

COULD ROGER FEDERER HAVE WON MORE GRAND SLAMS BY FOLLOWING A YOGA PROCESS?

Tristan Barnett [a,b]

[a] *Macquarie University*
[b] *Corresponding author: tristan.barnett@students.mq.edu.au*

Abstract
Roger Federer was winning a higher percentage of points on serve than his opponent but still losing Grand Slam matches due to not winning the important/pressure points and hence some degree of 'choking'. As identified in the academic literature and based on the author's experiences, by following yoga processes Federer would have most likely won at least 4 more slams and potentially up to 10 more slams; and taken Federer to a total of 'possibly' 24 slams to put him somewhat 'on par' in total Grand Slams with Djokovic and Nadal (24 and 22 slams respectively). Given that every player has a certain level of anxiety, every tennis coach could adopt these yoga strategies amongst their coaching regimes (where players could be 'close' to being free of all anxieties known as Samadhi or Enlightenment and hence 'very little' choking on critical stages of the matches). These yoga processes are outlined in the paper and to achieve results it has been identified that at least 7 years and possibly up to 11 years of yoga processes are required by an individual athlete. Hence, talent identification is required at a young age.

Keywords: Yoga, Bhagavad Gita, enlightened, anxieties

Acknowledgements
We wish to thank Professor Babette Pluim and Professor Ray Stefani for providing feedback on an earlier version of this manuscript.

References
Barnett T and Ejov V (2023). Does Roger Federer feel the pressure in close matches? *Science Meets Sports: when statistics are more than numbers.* Edited by Christophe Ley and Yves Dominicy.
Bedford A, Barnett T, Pollard GH and Pollard GN (2010). How the interpretation of match statistics affects player performance. *Journal of Medicine and Science in Tennis* 15(2), 23-27.
Bhagavad Gita As It Is (https://vedabase.io/en/library/bg/)
Kaur R (2021). Integrating yoga into athletic training: Enhancing performance and recovery in sports. *Journal of Sports Science and Nutrition* 2(2), 75-79
O'Leary TJ, Papadopulos MP and Van der Merwe A (2019). The impact of yoga-based core strength training on athletic performance and injury prevention. *J Sports Sci.* 37(8), 936-944
Searle A and Thomas G (2018). Yoga as a therapeutic approach: An exploration of its effectiveness in enhancing athletic performance. *J Sports Sci.* 36(7), 660-668.
Smith MS, Jones RL and Lyle J (2020). Yoga and athletic performance: A comprehensive review. *Sports Med.* 50(5), 915-928.
Srimad Bhagavatum (https://vedabase.io/en/library/sb/)

Submitted to MATHSPORT 2026

UNDERSTANDING HOW TODAY'S OLYMPIC SWIMMING CHAMPIONS MAINTAIN THE VELOCITY OF PAST CHAMPIONS WHILE SWIMMING TWICE AS FAR, SUPPORTING FUTURE IMPROVEMENTS

Raymond Stefani
Professor Emeritus, California State University Long Beach, USA
Raystefani@aol.com

Abstract

This paper examines what today's Olympic swimming champions have accomplished compared to the past, as a guide to what the future could hold. Winning times were taken for the seven men's and seven women's 2024 Olympic swimming races, which were also contested at half the distance. The number of years from 2024 to the year when half the 2024 winning time at the longer distance last won at half the distance, gives the years it took for the 2024 Olympic champion to swim at the same velocity as the past champion, but twice as far. The overall averages were 50 years for men and 46 years for women. For the 2024 200 m, 400 m and 800 m freestyle, women (52 years to 40 years) and men (60 years to 40 years) each took fewer years to double, at increasing distances. For the 200 m backstroke, breaststroke, butterfly and freestyle, the breaststroke took the fewest years for women (40) and men (44), while the other three stokes each took 52 years for women and took 56 years for men at 2 stokes and 60 years for the other stroke. The difference appears to be drag-affected underwater-motions for force application in the slower breaststroke verses above and below water leveraged motion for the faster strokes. For the 400m Individual Medley, women and men both took 40 years to double. Following the self-improvement meaning of Henri Didon's Olympic motto, *Citius, Altius, Fortius*, when young athletes get interested in swimming, they choose the methods of their day and try to improve, not being bound by past performances. Meanwhile, technicians improve methodologies shown by physics and kinesiology to improve velocity: nutrition, training, equipment, coaching and technique which, with the swimmer's philosophy of self-improvement, produces times that would have been thought unachievable 50 years ago.

Keywords: Swimming, Improved Performances, Physics, Kinesiology, Building Blocks Used for Improvement, Mind Set Used for Improvement,

1. INTRODUCTION

Olympic sports followers have recently marvelled at the performances during the Paris 2024 Olympics, as has been the natural reaction after each Games since the Modern Olympics began in 1896. Also, as has happened after each past Games, our curiosity was tweaked, causing us to wonder how much better the athletes will perform at the Los Angeles 2028 Olympics, now on our horizon, and then at future Games. That type of curiosity was even more tweaked at the Millenium, causing the sports network, ESPN, to air a program on 14 January 2000, in their *Outside the Lines* series, called *Bigger, Stronger, Faster: The Athlete of the 21st Century*, including four speakers, covering a wide spectrum of sports knowledge, ESPN (2000). We will use one extraordinarily insightful statement as the template for this paper. Many of the ideas presented at ESPN (2000) will also be used.

Robert Newton, an expert on biomechanics at Ball State University, discussed devices which are used to measure performance, to strengthen athletes, and to create new and better techniques. Bob Kersee covered coaching methods in depth. His wife, Jackie Joyner-Kursee, had won 6 medals in Olympic athletics. He covered the many diverse physical and psychological skills a coach can teach athletes and insightful methods to enhance those skills. Michael Johnson, the then world record holder at the 200 m and 400 m runs, discussed an athlete's perspective in depth: training, competing, improving and dealing with the psychological stresses of competition. The fourth speaker was Raymond Stefani, me, who had just contributed a featured article, *Faster, Higher, Farther*, to the sports section of the Millenium edition of the NY Times. I covered record setting in depth, looking at various sports, various events, how men and women have improved relatively, and that we can expect records to continue to be broken, even Michael Johnson's.

Summing up, the best advice about how to look at athletic future accomplishments came from Michael Johnson, who said: "No one knows how much is left. We only know where we've been." Following that

observation, it makes sense to focus this paper on what today's athletes have accomplished compared to the past. Good success in the past can reassure us that there will likely be future progress.

A remarkable improvement in the past began in 1924, before which no men's Olympic swimming champion had ever completed 100 m in under 60 s. Typically, shorter distance sprint events records are broken in tenths of a second. US swimmer Johnny Weissmuller astounded the sports world by winning in 59 s, a full second below the previous record. That was so widely acclaimed that Weissmuller was chosen to act as Tarzan in the movies. His credibility at playing the role of such a strong character was created by having set such an astounding record. A 1924 observer would have doubted that anyone could maintain Weissmuller's velocity much farther than 100 m. Let us move 80 years later, when Australian swimmer Grant Hackett won the 1500 m, 15 times 100 m, in a time of 14:43.40, 1.6 seconds faster than 15 times 59 s. Figuratively, Hackett swimming continuously would have beaten a relay of 15 Weissmullers by 1.6 s, a feat that would not have been remotely believable in 1924.

That is an amazing improvement, but only one. For a meaningful study, we need a much larger sample. In Section 2, the winning time is taken for all 14 2024 Olympic swimming champions for the races also contested at half the distance. We proceed by dividing each 2024 winning time at the longer distance by 2 and by finding when that time last won at half the distance of the 2024 race. The velocity would be the same. The number of years between each former and current Olympic champion is the number of years it took for the current champion to swim twice as far at the same velocity, which we will abbreviate **TFSV** henceforth. The 7 men's and 7 women's results will be compared for distance and gender affects.

In Section 3, physics and kinesiology will be used to find the velocity ratio for swimmers, which can then be analysed to identify the various building blocks that must be improved to swim TFSV. In Section 4, we connect the results from Sections 2 and 3 to identify an athlete's mind set, allowing the application of the improved building blocks to achieve TFSV. The Olympic Motto and Creed will prove explanatory. Section 5 summarizes conclusions.

2. GOING TWICE AS FAR AT THE SAME VELOCITY (TFSV)

As mentioned above, we divided the 2024 Olympic winning time by 2 for each of the 14 events (7 for man and 7 for women) also contested at half the distance. By finding when that time last won at half the distance, the number of years between each former and current Olympic champion is the number of years it took for the current champion to swim twice as far at the same velocity (TFSV). The results are summarized in Table 1. World Almanac (2026) gives all Olympic champions and their winning performances. Walechinsky, D. and Louky, L. (2012) provide the first six placers in each event with a commentary on each sport and each competition.

Stroke	**2024 Distance**	**Earlier Half Distance**	**Years for Men to Go Twice as Far at the Same Velocity**	**Years for Women to Go Twice as Far at the Same Velocity**
Freestyle	200 m	100 m	**60**	**52**
	400 m	200 m	**52**	**48**
	800 m	400 m	**40**	**40**
Back	200 m	100 m	**56**	**52**
Breast	200 m	100 m	**44**	**40**
Butterfly	200 m	100 m	**56**	**52**
IM	400 m	200 m	**40**	**40**
Average			**49.7**	**46.3**

Table 1: Going twice as far at the same velocity for 14 2024 Olympic Champions

The first three rows of Table 1 refer to the same stroke, freestyle, at different 2024 winning distances (200 m, 400 m and 800 m). The number of years to achieve TFSV was progressively lower for both women (from 52 to 40) and men (from 60 to 40) at progressively longer distances. Women took fewer years than men for the 200 m and 400 m, and the same 40 years for 800 m. Progressively longer distances mean progressively lower velocities and champion women win in more time (less velocity) than champion men. Thus, lower velocity is consistent with fewer years to achieve TFSV.

Table 1 includes 2024 data for all four different stokes at the same 2024 200 m distance for men and women. Women took their least TFSV years for the breaststroke with 40 years and the exact same 52 years for

the other three strokes. Men also took their least time, 44 years, for the breaststroke, 56 years for both the backstroke and butterfly, while freestyle took the smallest possible additional increase of one Olympiad, taking 60 years. Cleary, two groups emerge, breaststroke years for TFSV and the years for the other three strokes TFSV. A similar division is the drag-affected underwater-motion for force application in the slower breaststroke, verses above and below water leveraged motion for the three faster strokes. Women took fewer years than men to achieve TFSV in all four strokes at 200 m, consistent with winning by lower velocities at the same distance.

For the 400m Individual Medley, women and men both took 40 years to achieve TFSV, consistent with averaging across all four strokes and with the longer distance creating the same gender-related TFSV as was the case for the 800 m freestyle. The average years for men to TFSV was 49.7, compared to 46.3 years for women.

While Table 1 provides much insight into the causes of the differing number of years taken by gender, stroke and distance. One overarching observation remains: 14 2024 male and female Olympic champions swam twice as far at the same velocity as did their half-distance counterparts about two generations earlier, which would have been considered most unlikely if not impossible by those predecessors and by the observers of that time.

Next, we identify the building blocks that can make improvements possible.

3. BUILDING BLOCKS FOR INCREASED VELOCIY

To identify the building blocks for improvement, it is necessary to first evaluate the dynamics of swimming in a form that facilitates such evaluation. Stefani (2002, 2007, 2020) examines references covering the physics and kinesiology of swimming, resulting in equations for swimming velocity. The power the swimmer applies to push against the water is identified in (1), written mostly without algebraic characters, where LBM means Lean Body Mass and m means body mass.

$$\text{LBM x training x m x swimming efficiency} \tag{1}$$

The power the water pulls back on the swimmer due to drag is given in readable form in (2)

$$\text{Velocity}^3 \text{ x surface area } (m^{2/3}) \text{ x the drag coeff. x constants} \tag{2}$$

Next, (1) and (2) are equated and solved for velocity, v, shown in (3), using algebraic characters. (3) provides the ratio of two velocities, perhaps by two different individuals or by the same individual at two different times. Constants cancel.

$$v_2/v_1 = [(Tr_2/Tr_2)\,(e_2/e_1)]^{1/3}\,(LTW_2/LTW_1) \,/\, (Cd_2/Cd_1)]^{1/3}\,(m_2/m_1)^{4/9} \tag{3}$$

The right side of (3), from left to right, gives ratios for training, efficiency, lean to weight, drag coefficient and body mass. (3) is too complicated to provide a simple means to identify building blocks of performance. Stefani (2014a, 2014b) contains a thorough evaluation of 10 references including data for 2286 elite athletes. Several tables are created, by which it follows in (4), that lean to weight ratio is approximately equal to drag ratio for elite athletes and in (5) that the square of the lean to weight ratio is approximately equal to body mass ratio.

$$(LTW_2/LTW_1) \approx (Cd_2/Cd_1) \tag{4}$$

$$(LTW_2/LTW_1)^{\,2} \approx (m_2/m_1) \tag{5}$$

(4) and (5) can be used to simplify (3) with (6), by which it becomes much easier to identify building blocks for performance.

$$v_2/v_1 \approx [\,(LTW_2/LTW_1)\,]^{\,8/9}\,[\,(Tr_2/Tr_1)\,(e_2/e_1)\,]^{1/3} \tag{6}$$

In (6), we see that the velocity ratio depends on ratios of lean to weight, training and efficiency. An athlete begins with some starting lean to weight value and then can increase that value, using **nutrition** as a building block. Well-chosen and effective **equipment** may be used to improve training outside of competition or to be used during competition to improve techniques by which the athlete's power can be more efficiently converted

into performance. The athlete's lean to weight value, as affected by **training,** produces power applicable to performing. **Coaching** can improve both training and efficiency. Effective **technique** can improve efficiency. Power times efficiency gives performance. The building blocks that can improve performance, are thus **nutrition, equipment, training, coaching and technique.** Table 2 demonstrates how four applications of new building blocks dramatically improved performance.

Breakthrough	Building Block	Average %I/0 1956 to Before Introduction	%I/O on First Use (Year)	% Increase in %I/O
Rowing Machine	Equipment for Training	1.22	7.42 (1980)	**508**
Fiberglass Pole	Equipment for Technique	1.64	8.51 (1964)	**419**
Fosbury Flop	Technique	2.25	4.12 (1968 Men) (1972 Women)	**83**
Clap Skate	Equipment for Technique	1.82	2.88 (1998)	**58**

Table 2: Percent Improvement Per Olympiad (%I/O) Enhanced by Building Blocks

For each of the four building block examples in Table 2, the average percent improvement per Olympiad (%I/O) was calculated for that sport's Olympic champions from 1956, when the effects of WW2 had ended, to the last competition prior to the application of the new building block. The percent improvement at first use is then compared to the average prior to use to identify the percent change created by introduction of each new building block.

Before the rowing machine was introduced prior to the 1980 Olympics as equipment used for training, the ability of a rower to train effectively was limited by weather, water conditions, availability of boats and work time demands. When the rowing machine became available, the training athlete could pull on the cords attached to a load demanding wheel, closely simulated the muscle-specific loading with body motion required by actual rowing, but on a streamlined machine available at any time, Uddin (2025). The greatly increased opportunity for effective training increased the %I/O by a very remarkable 508%.

Before the fiberglass pole became available for the 1964 Games as an equipment-enhancing technique, a pole vaulter had to run at a very careful velocity to put the stiff bamboo or wood pole into a slot, hold onto the stiff pole as the athlete and pole swung forward and them try to clear the bar. Instead, given the very flexible fiberglass pole, the athlete could approach the slot much faster, insert the pole, and continue to move forward and upward as the pole flexed, Track Coach (2026). The pole then straightened, swinging the vaulter upward. By carefully coordinated action, the vaulter then cleared the bar. On first use, %I/O increased by a significant 419%. The winning heights had stayed close to 15 ft for several years, but that height has now reached 20 ft.

Before the first Olympics use of the Fosbury Flop technique by men in 1968 and by women in 1972, the earliest high jumpers simply vaulted the bar. Later and more successfully, high jumpers either straddled the bar horizontally, with face and chest downward, Wikipedia Straddle Technique (2026); or rolled horizontally around the bar with their side or back facing the bar as they rotated, Wikipedia Western Roll (2026). I have reviewed photographs of the use of the Straddle and Western Roll, and I estimate that the athlete's centre of gravity had to be about one foot over the bar to clear it. The new Fosbury Flop method was to push upward head-first, rotating on the back to clear the bar at 90 degrees, resulting with the bar being only slightly below the jumper's centre of gravity, Wikipedia Fosbury Flop (2026). The distance that the bar was below the jumper's centre of gravity was reduced, which is why %I/O increased by 83% on the first use of the Fosbury Flop.

Before the first use of the clap skate in 1998 for Olympic speed skating as equipment for enhancing technique, when the skater pushed forward with skate blades firmly attached, only the tip of the rear blade would touch the ice. Instead, the clap skate blades were hinged so when pushing forward the rear blade would rotate, remaining in contact with the ice, then snap back to the skate making a clapping sound, Wikipedia Clap Skate (2026). The greater efficiency of pushing against the ice caused the %I/O to increase 58% on first use.

These four innovations were particularly successful across the spectrum of sports. Many other improvements in building blocks have not had as dramatic an effect on %I/O, but in general, improvements in building blocks along with athlete improvements have continued to increase %I/O in sports in general and in swimming in particular, as noted in Table 1.

Knowing the enormous improvements of Section 2 and knowing the available building blocks of performance in Section 3, which themselves improve in time, Section 4 explores the mind set by which an athlete can employ those building blocks and proceed into competitive success.

4. THE MIND SET BY WHICH SWIMMERS CAN ACHIEVE THEIR RESULTS

On 7 March 1891, at an Arcueil College sports assembly in France, Father Henri Didon (1840-1900), a Dominican Priest, said that to be a success in sport, a person should improve little by little, that is, to run a bit faster, to jump a bit higher and to become a bit stronger (*Citius, Altius, Fortius* in Latin) than they had done before, IOC (2021). He didn't say anything about beating anyone or setting a record. He gave them a motto in sport and in life: *Citius, Altius, Fortius*, encouraging self-improvement as a satisfying goal. He did not require competition in sports, but that was suggested.

His friend, Baron de Coubertin, was present for Didon's speech in 1891. When de Coubertin founded the modern Olympic movement in 1894, he chose *Citius, Altius, Fortius* as the Olympic Motto. Baron de Coubertin decided that it was important to clarify the meaning of the Olympic Motto, by creating the Olympic Creed, based on a talk given at the 1908 Olympics by Ethelbert Talbot, Bishop of Pennsylvania. Coubertin's Olympic Creed reads "The most important thing in the Olympic Games is not to win but to take part, just as the most important thing in life is not the triumph but the struggle. The essential thing is not to have conquered but to have fought well."

Following the philosophy of the Olympic Motto and Creed, when a young person gets interested in physical activity, they choose the methods of that day and age, not thinking about the past or wondering if anyone in the future will do better. The vast majority are just having some fun with physical activity and sports competition, in a satisfying and constructive manner. While these young athletes are doing their best, others are working to improve the building blocks identified above: nutrition, equipment, training, coaching and technique. Those improvements to the building blocks elevate what the athlete can achieve, collectively creating the remarkable accomplishments of Table 1, because the best of these improving athletes will be on sports teams using their building blocks, the best of them will be on national teams, and the best of them will win Olympic gold medals.

PERSONAL OBSERVATIONS

In Section 1, I mentioned that I personally heard an opinion about the importance of looking at past accomplishments voiced by then world record holder Michael Johnson on a millennium ESPN program that I took part in. I also realize that in my own sports career and those I witnessed, the building blocks and mind sets mentioned above were prominent.

As a competitive swimmer in high school, I went day to day, trying to do my best, effectively following Henri Didon's *Citiu*s and available building blocks to be successfully competitive at the 200 yd freestyle. When I began swimming in college, I was not eligible to compete in my first year, but I could train and observe. I observed two seniors contributing to team success as competitors at 440 yd, more than twice my longest previous distance. I was highly impressed that anyone could compete at that distance. A year later, after they had graduated, my coach asked me to begin training to swim at 440 yd, aided by a teammate who had swum that distance. I basically followed the Olympic Motto and Creed, trying to increase speed, struggle and fight as well as I could. I learned much from the demands of the first swimming meets and used available building blocks. After 5 meets, I had taken 10 s off the school record. I did not feel superiority over those whose records I had broken, but instead I felt gratitude for their contributions, which I told them. Toward the end of my college career, I again followed the Olympic Creed, and I decided to struggle and fight well to complete a 1500 m time trial, which met the US Olympic trials qualifying standard.

After I reached graduate school, I worked to satisfy the tests of a national sport honorary society. I realized that one athlete seemed to live the Creed "the most important thing in life is not the triumph but the struggle". He rigorously and successfully took on each sports challenge and seemed to enjoy each moment, as he completed every requirement with much to spare. His clear versatility led coaches and college officials to raise money to send him to the Maccabiah Games, nicknamed the Jewish Olympics. He not only competed, but he was the Maccabiah Games champion of the decathlon, a result of his versatility and willingness to take on new challenges and to improve step by step, the essence of the Olympic Motto and the Olympic Creed.

I realize I have seen many examples of the use of building blocks and how success was achieved by an adventuresome and open-minded view of sports. The advice I got directly from one of the greatest athletes and honourable people of all time also comes to mind, as discussed below.

THE BEST ADVICE CAME FROM AN ALL TIME BEST, **JESSE OWENS**

In about 1957, my predominantly white Detroit west-side high school swimming team was scheduled to compete with a predominantly black east-side high school, but neither school had a pool. Both schools usually swam all meets away. The Detroit organizers found a school pool in the downtown for us both to compete at. The story was in the local newspapers as a case of racial equity.

When both teams met at the tiny, designated pool, the meet official welcomed us and told us a very special person wanted to speak to us, Jesse Owens, who had read about our meet in the newspaper while on business in Detroit. We were mesmerized to say the least. Jesse spoke in a very open, accommodating, friendly and inspiring way. He discussed how to handle adversity. He summarized by saying **"There is nothing you can do about what people say about you or think about you. There is one thing you can always do. You can always do your best."** We all tried to swim our career bests: not for ourselves, not for our teams, but for Jesse. One of my greatest memories was winning twice and shaking hands with Jesse Owens twice. Years later, after learning more about Jesse's life, I realized that what he said was his life's theme.

His words should be not just a mantra for sport, but also for life.

5. CONCLUSIONS

Following the advice given by then world record holder Michael Johnson at the millennium, "No one knows how much is left. We only know where we've been.", we looked to past success and found that all 7 2024 male and all 7 2024 female Olympic champions swam twice as far at the same velocity as did their half-distance counterparts about two generations earlier, which would have been considered most unlikely if not impossible by those predecessors and by the observers of that time. To achieve that remarkable progress, the application of physics and kinesiology identifies the building blocks that can improve swimming velocity: nutrition, equipment, training, coaching and technique.

Following the philosophy of the Olympic Motto and Creed, when a young person gets interested in physical activity, they choose the methods of that day and age, and do their best to improve, not thinking about the past. While these young athletes are doing their best, others are working to improve the building blocks identified above: nutrition, equipment, training, coaching and technique. Those improvements to the building blocks elevate what the athlete can achieve, collectively creating the remarkable accomplishments of going twice as far as former champions ate the same velocity.

I was present when Jesse Owens said **"There is nothing you can do about what people say about you or think about you. There is one thing you can always do. You can always do your best."** That approach over two generations could create the great improvements and is also what ought to be our manta for life in general.

References

IOC (2021) The Olympic Motto, 20 July 2021, https://www.olympics.com/ioc/olympic-motto

ESPN (2000) Bigger, Stronger, Faster: The Athlete of the 21st Century, *ESPN Series Outside the Lines, Episode 11, Number 1, Jan 14, 2000*.

Stefani, R.T (2002) The Fundamental Nature of Differential Male/Female World and Olympic Winning Performances, Sports and Rating Systems, *Proceedings of the Sixth Australasian Conference on Mathematics and Computers in Sport, July 1-3, 2002, pp 32-46*.

Stefani, R.T. (2007) The Physics and Evolution of Olympic Winning Performances, Chapter 3, *Statistical Thinking in Sports. Chapman and Hall/CRC Press: Boca Raton, London, New York*.

Stefani, R.T. (2014a) The Power-to weight relationships and efficiency improvements of Olympic champions in athletics, swimming and rowing, *International Journal of Sports Science and Coaching, 9(2)*.

Stefani, R.T. (2014b) Understanding the Velocity Ratio of Male and Female Olympic Champions in Running, Speed Skating, Rowing and Swimming, *Proceedings of the 12th Australasian Conference on Mathematics and Computers in Sport, Darwin Australia, 25-27 June, 2014. pp 106-111*.

Stefani, R.T (2020) Understanding the Female/Male Velocity Ratio Versus Distance for All Swimming Strokes, Including English and Catalina Channel Swims, *Proceedings, 15th Australasian Conference on Mathematics and Computers in Sport (ANZIAM Mathsport 2020) Victoria University, Wellington, New Zealand, 9-11 November 2020, pp 106-113*.

Track Coach (2020) A Practical Guide to Fiber Glass Pole Vaulting, *Track and Field News, 6 January 2020*, https://trackandfieldnews.com/track-coach/a-practical-guide-to-fiberglass-pole-vaulting/ .

Uddin, Jahir (2025) What Does a Rowing Machine Work? A Complete Guide, *Linkedin, 25 August 2025*, https://www.linkedin.com/pulse/what-does-rowing-machine-work-complete-guide-jahir-uddin--bcyyc .
Walechinsky, D. and Louky, L. (2012) The Complete Book of the Olympics 2012 Edition, Aurum Press Ltd, London, ISBN 978-1-84513-695-6.
Wikipedia Clap Skate (2026) Clap skate, https://en.wikipedia.org/wiki/Clap_skate , retrieved 1 March 2026.
Wikipedia Fosbury Flop (2026) Fosbury flop, https://en.wikipedia.org/wiki/Fosbury_flop , retrieved 1 March 2026.
Wikipedia Straddle Technique (2026). Straddle technique, https://en.wikipedia.org/wiki/Straddle_technique, retrieved 1 March 2026.
Wikipedia Western Roll (2026) Western roll, https://en.wikipedia.org/wiki/Western_roll , retrieved 1 March 2026.
World Almanac (2026). The World Almanac and Book of Facts 2026, Skyhorse Publishing, USA, ISBN-13, 978-1-5107-8398-0.

.