A Transparent Point-level Momentum Framework for Professional Tennis: Evidence from Wimbledon 2023

Jiuwei Huang*
Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia
*Corresponding email: jiuwei.lab@gmail.com

Abstract

Momentum is frequently invoked to explain abrupt shifts in professional tennis, yet many quantitative descriptions either depend on opaque machine-learning outputs or leak post-point information into predictive claims. This study develops a transparent, point-level momentum framework using the public 2023 Wimbledon Gentlemen's singles point-by-point dataset released with the 2024 Mathematical Contest in Modeling Problem C. The dataset contains 7,284 point records from 31 matches after the first two rounds. We transform point events into signed player-relative increments, combine them through an exponentially weighted moving average, and then evaluate whether the resulting momentum score adds information beyond pre-point score context and serving status. The reconstructed framework separates descriptive visualization from prediction: The momentum index is used to describe match flow, while next-point prediction is evaluated only with information available before a point begins. Results show that serving status remains the strongest predictor of the next point. A score-and-serve baseline achieved a group cross-validation area under the curve (AUC) of 0.683; adding exponentially weighted moving average (EWMA) momentum raised AUC to 0.687, and a full contextual model reached 0.690. This indicates that momentum contains limited but non-negligible incremental signal rather than near-perfect determinism. Large momentum swings were concentrated around winners, aces, close-to-line serves, successful net points, and double faults. The proposed framework provides a reproducible alternative to overfitted momentum models and offers coaches a practical dashboard for identifying high-leverage tactical events without claiming causal psychological measurement.

Keywords

Tennis analytics, Psychological momentum, Wimbledon, Point-by-point data, Exponentially weighted moving average, Machine learning

Introduction

In professional tennis, the word momentum is often used when a player appears to gain temporary control after a sequence of successful shots, a break of serve, or a high-pressure point. In the sports-psychology literature, momentum has long been described as a multidimensional process involving changes in cognition, affect, physiology, and performance. Recent tennis-specific work has likewise shown that match observers often perceive short-run directional control even when the underlying process remains noisy and probabilistic [1]. The 2023 Wimbledon Gentlemen's singles final between Carlos Alcaraz and Novak Djokovic made this problem especially visible because control of the match shifted across sets and within games. These visible swings create a practical question for coaches: Can match momentum be quantified in a way that is useful for tactical review without exaggerating its predictive power?

The earlier contest-style version of this work treated momentum as a black-box modeling task. That structure is not ideal for journal submission because it mixed descriptive fitting, outcome prediction, and advice in the same pipeline. It also resembled recent data-driven tennis studies that emphasize algorithmic prediction, including random-forest and boosting style approaches, but often make it difficult to separate signal from leakage or post-point information [2]. A publishable manuscript requires a clearer distinction between what is being measured, what is being predicted, and which variables are available at each time step.

The present revision therefore reframes the work as a transparent sports-analytics study. We do not attempt to measure a hidden psychological state directly. Instead, we define an operational point-level momentum score that summarizes recent observable events while explicitly separating descriptive visualization from pre-point prediction. This design is consistent with prior cautions. Perceived momentum in tennis can be confounded by serve structure, score pressure, and strategic sequencing unless the analyst controls what information is available when a prediction is made.

This study addresses four questions. First, how can point-by-point tennis events be converted into a player-relative momentum score? Second, what does this score reveal about the flow of the 2023 Wimbledon final and the surrounding tournament matches? Third, does the momentum score improve next-point prediction beyond score state and serving advantage? Fourth, which observable events are most associated with large momentum swings and therefore most actionable for players and coaches?

Literature background

Momentum in sport has both psychological and strategic meanings. Momentum has been described as a multidimensional process in which precipitating events alter cognition, affect, physiological readiness, and subsequent performance. In racket sports, this concept is attractive because players and spectators regularly observe bursts of apparently one-sided control. At the same time, empirical work has warned that such perceived streaks can be amplified by selective attention and by the structure of the game itself. So momentum should be studied cautiously rather than treated as an unquestioned causal force [3].

Tennis is a demanding case because service advantage, score pressure, and alternating game structure already create strong serial patterns. Previous work has shown how point-level tennis data can support probabilistic forecasting, while later work demonstrated that observed runs and reversals may blend strategic momentum with psychological momentum. Given these challenges, this means that any useful momentum indicator must remain interpretable, must respect match chronology, and must not quietly borrow information from the future [4].

Machine-learning studies have recently revisited tennis momentum using the same Wimbledon 2023 data source, including Categorical Boosting (CatBoost), random forests, Extreme Gradient Boosting (XGBoost), and neural networks. These approaches are useful benchmarks because they can detect nonlinear interactions, and ensemble learners in particular are often strong predictors in structured tabular settings. However, if post-point variables are fed into models that are then interpreted as if they were pre-point forecasts, apparent accuracy can be overstated. The present study therefore prioritizes transparency, temporal validity, and figure-level interpretability over black-box performance alone [5].

Data and methods

Data source and scope

The primary data source is the point-by-point dataset released with the Consortium for Mathematics and its Applications (COMAP) 2024 Mathematical Contest in Modeling, Problem C, Momentum in Tennis. For computation, the public featured-match file was cross-checked against the associated public GitHub repository that mirrors the competition materials. The dataset contains 7,284 point records spanning 31 matches after the first two rounds, including the Wimbledon 2023 Gentlemen's singles tournament from the Round of 32 through the final, as summarized in Table 1 and 2.

Table 1. Figure-data mapping.

Figure	Data used	Generation method
Figure 1	Method workflow and manuscript design logic	Code-drawn schematic based on the revised analytical workflow; no empirical observations are plotted.
Figure 2	Wimbledon_featured_matches.csv	Grouped by match_id and inferred tournament round from match_id.
Figure 3	match_id = 2023-wimbledon-1701	EWMA momentum score computed directly from the final-match point sequence (match_id = 2023-wimbledon-1701), with set boundaries and top decile one-step swings marked.

Figure	Data used	Generation method
Figure 4	All 31 matches	All 31 match momentum series interpolated to 100 normalized progress bins, summarized by a tournament-level median/IQR panel and a round-stratified heatmap ordered by average absolute momentum.
Figure 5	All 31 matches	Five-fold group cross-validation by match_id; logistic regression with standardized predictors.
Figure 6	All 31 matches	Average absolute contribution of each component in Δt.
Figure 7	All 31 matches	Event rates on top 10.0% absolute one-step momentum changes versus all other points.
Figure 8	All 31 matches	Transition probabilities among P2 edge, neutral, and P1 edge states using thresholds of ±0.65 SD.

Table 2. Point-performance increment components.

Component	Definition	Weight
Point won	+1 if Player 1 wins the point; -1 if Player 2 wins the point	1.00
Ace/winner	Player 1 ace/winner minus Player 2 ace/winner	0.35 each
Errors avoided	Opponent unforced error or double fault favors the focal player	0.45 for unforced error; 0.60 for double fault
Break point	Break-point conversion favors converter; missed break point penalizes challenger	0.80 conversion; -0.35 miss
Net success	Net point won by Player 1 minus net point won by Player 2	0.25
Serve-return depth	Deep return by returner and close-to-line serve by server, signed by player	0.20 return; 0.15 serve
Speed/movement	Within-match standardized serve speed signed by server and movement balance	0.10 each

The variables include match identity, player names, elapsed time, set, game and point numbers, score state, server, serve number, point winner, cumulative points, game and set victors, as well as aces, winners, double faults, unforced errors, net points, break points, running distance, rally count, serve speed, serve width, serve depth, and return depth. Missing values occur primarily in return_depth (1,309 records), speed_mph (752 records), serve_width (54 records), and serve_depth (54 records). These missingness patterns are reported instead of hidden, because serve and return variables are central to tennis analytics.

No private player information, biometric information, betting data, or weather data were used. The analysis is therefore limited to observable point-level technical and scoring variables. This limitation is important: The results describe statistical match flow, not the full mental or physical condition of the players [6].

Table 3. Dataset summary and transparency checks.

Item	Value	Notes
Raw point records	7,284	Rows in Wimbledon_featured_matches.csv
Matches	31	Round 3 through final, identified by match_id
Players	32	Union of player1 and player2 fields
Most missing field	return_depth	1,309 missing records
Other missing fields	speed_mph, serve_width, serve_depth	752, 54, and 54 missing records, respectively
Unit of analysis	Point	Each row represents one played point

Preprocessing

Elapsed time was converted from H:MM:SS to seconds. Match order was preserved using the original row order within each match_id. Categorical variables were not blindly one-hot encoded for all downstream tasks; instead, only variables with direct tactical meaning were recoded for the relevant analyses.

For example, server was converted to a signed indicator from the perspective of Player 1, while serve_depth and return_depth were converted into player-relative tactical advantages.

Serve speed was standardized within each match to avoid comparing raw speeds across players and match contexts. Running-distance balance was also standardized within match as the difference between the opponent's distance and the focal player's distance. Missing serve speeds were treated as zero after within-match standardization only for the contribution term, meaning that missing speeds do not create artificial speed advantages. Missing return depth was treated as no deep-return contribution rather than imputed as deep or not deep.

The target variable for pre-point prediction is whether Player 1 wins the current point. Only information known before the point begins, such as serving status, score state, break-point state, and lagged momentum, is used in such models. This timing rule follows a simple but critical principle from predictive sports analytics: the model should not be allowed to see consequences of the point it is asked to predict [7].

Operational momentum score

Let Δt denote the signed point-performance increment after point t. Positive values favor Player 1 and negative values favor Player 2. The increment combines observable events using transparent weights chosen to reflect tennis-specific leverage. Winning the point receives the largest contribution. Aces, winners, errors, double faults, break points, net-point success, serve/return depth, serve speed, and movement balance provide smaller modifiers. The exact components are reported in the reproducibility appendix.

Momentum is then computed, as shown in Figure 1, as an exponentially weighted moving average: $Mt = (1 - \alpha)Mt-1 + \alpha\Delta t$, where $\alpha = 0.18$. This choice gives the score memory over several recent points while still allowing quick response to high-leverage events. The weighting scheme is intended to be transparent rather than optimal in a machine-learning sense, which keeps the resulting score interpretable for match review and tactical diagnosis.

A momentum state is defined using thresholds of ±0.65 standard deviations: P1 edge, neutral, and P2 edge. A large swing point is defined as a point whose absolute one-step change in momentum is in the top 10.0% of all point-to-point changes. The ±0.65 threshold is used as a practical visualization cut-off and is broadly consistent with the idea of a medium standardized effect region in behavioral interpretation. These thresholds are intended to support visualization and coaching interpretation; they are not claimed to be universal constants [8].

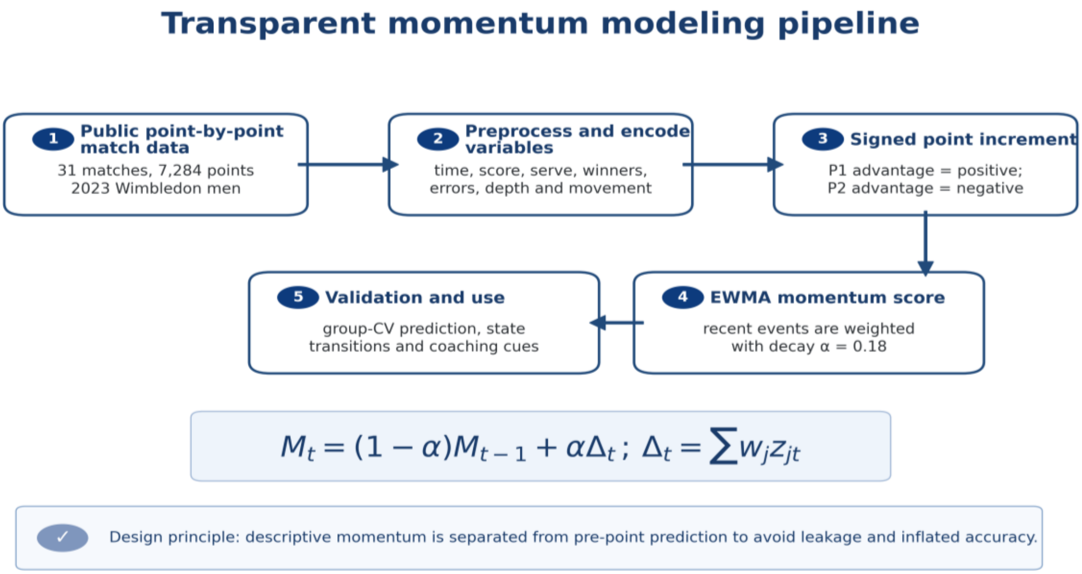

Figure 1. Code-drawn momentum-modeling pipeline (The descriptive momentum score is intentionally separated from pre-point prediction to reduce leakage and overfitting).

Statistical and predictive evaluation

The evaluation is conservative. We estimate whether lagged momentum improves prediction of the next point beyond a baseline consisting of score state and serving status. Models are evaluated by five-fold group cross-validation with match_id grouping, so no same-match points appear in both training and testing folds. This design gives a more realistic generalization estimate than random splits, due to within-match dependence is substantial.

Logistic regression with standardized predictors is used as the main validation model because its coefficients remain interpretable. Area under the receiver operating characteristic (ROC) curve (i.e, AUC), accuracy, and Brier score are reported following common predictive-performance practice. Random forests were tested for exploratory diagnostics. However, the manuscript emphasizes transparent coefficients and event contributions rather than black-box accuracy. All computation was implemented in a standard Python/ scikit-learn workflow.

Results

Tournament data structure

Figure 2 summarizes the data coverage and match-length structure. The dataset is balanced by tournament structure rather than by point count: Earlier included rounds have more matches and therefore more point records. In the revised left panel, point totals and match counts are annotated directly above each bar, so no secondary y-axis is needed. The right panel orders all matches by record length and highlights the final, making the variation in match length clear without crowding the y-axis with every match_id [9].

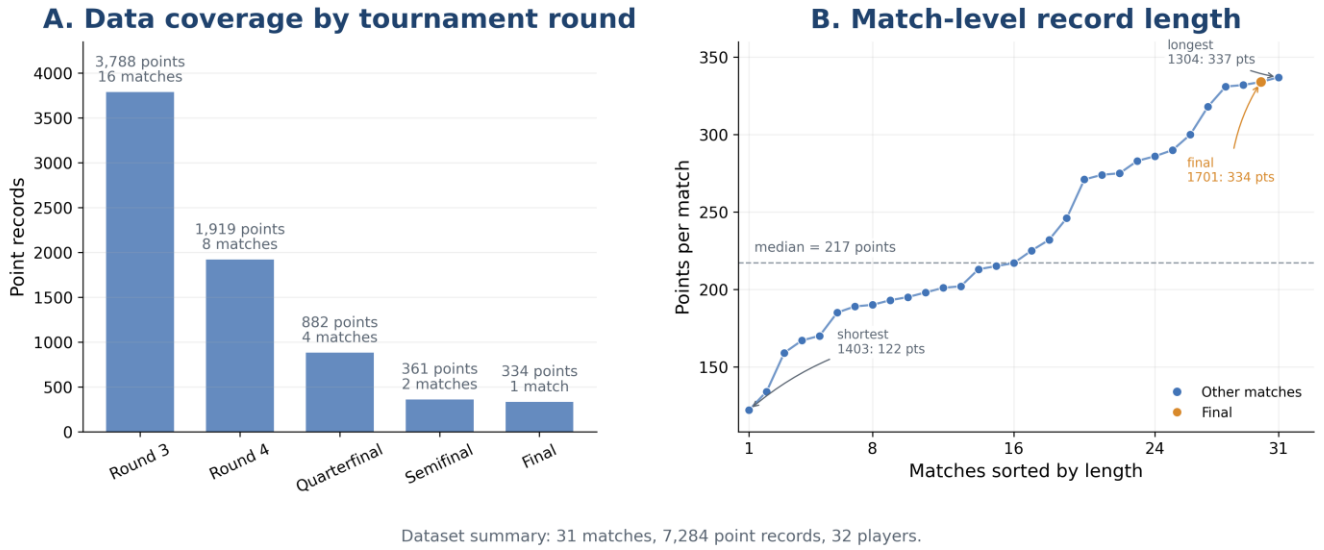

Figure 2. Data atlas (Panel A shows point records by tournament round, with point totals and match counts annotated directly above each bar. Panel B orders the 31 matches by record length, highlights the final, and marks the median match length. The centered note below the panels summarizes the overall dataset size).

Momentum flow in the Wimbledon final

Figure 3 applies the momentum score to the final (match_id 2023-wimbledon-1701). The light line reports raw point-level momentum and the bold line reports the EWMA trend. The pattern is not a monotonic climb by one player; instead, directional control alternates across sets. Set 1 favors Djokovic (mean = -0.89), Sets 2 and 3 shift toward Alcaraz (0.13 and 0.51), Set 4 tilts back toward Djokovic (-0.27), and Set 5 again leans toward Alcaraz (0.16). The clearest sustained positive regime occurs late in Set 3 and again early in Set 5.

The orange markers identify the largest one-step changes in standardized momentum. These points cluster around local turning points rather than merely around the final scoreline, which is why the figure is useful as a match-diagnostics plot. The combined raw and smoothed views support interpreting momentum as short-run directional control rather than as a deterministic hidden force. This diagnostic perspective allows coaches and analysts to identify exactly when and where matches are decided, moving beyond final statistics to capture the dynamics that shape critical moments.

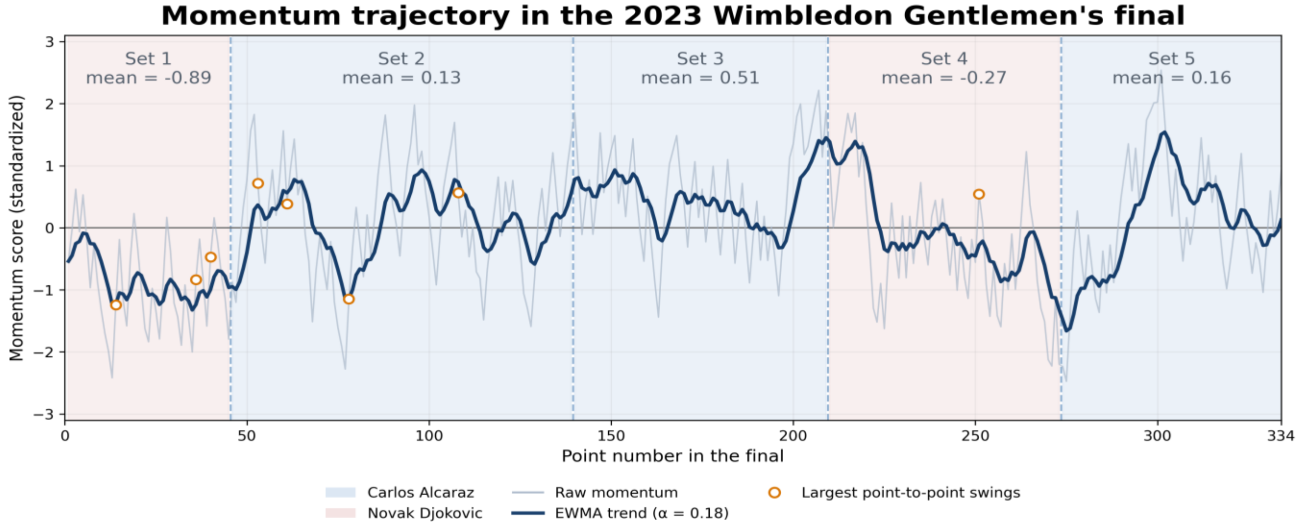

Figure 3. Code-generated momentum trajectory in the 2023 Wimbledon Gentlemen's final (The light line is raw point-level momentum and the bold line is the EWMA trend; shaded set blocks indicate the set winner, and dashed lines indicate set boundaries. Positive values favor Carlos Alcaraz and negative values favor Novak Djokovic).

Cross-match momentum atlas

Figure 4 further generalizes the final-match view to all 31 matches Each match-specific momentum sequence is linearly interpolated to a common 0.0-100.0% progress scale, and rows are ordered within each round by average absolute momentum. The top panel also reports the tournament-wide median trajectory with an interquartile ribbon, whereas the bottom heatmap shows between-match heterogeneity after within-match standardization.

Two patterns are clearer in this representation than in the earlier single heatmap. First, the tournament-level median momentum remains close to zero for much of the first 60.0% of normalized match time, then turns modestly positive between roughly 65.0% and 80.0% before moving back toward balance. Second, later rounds display broader and more persistent color bands, indicating that semifinal and final matches contain longer contiguous runs of directional control than many earlier-round matches.

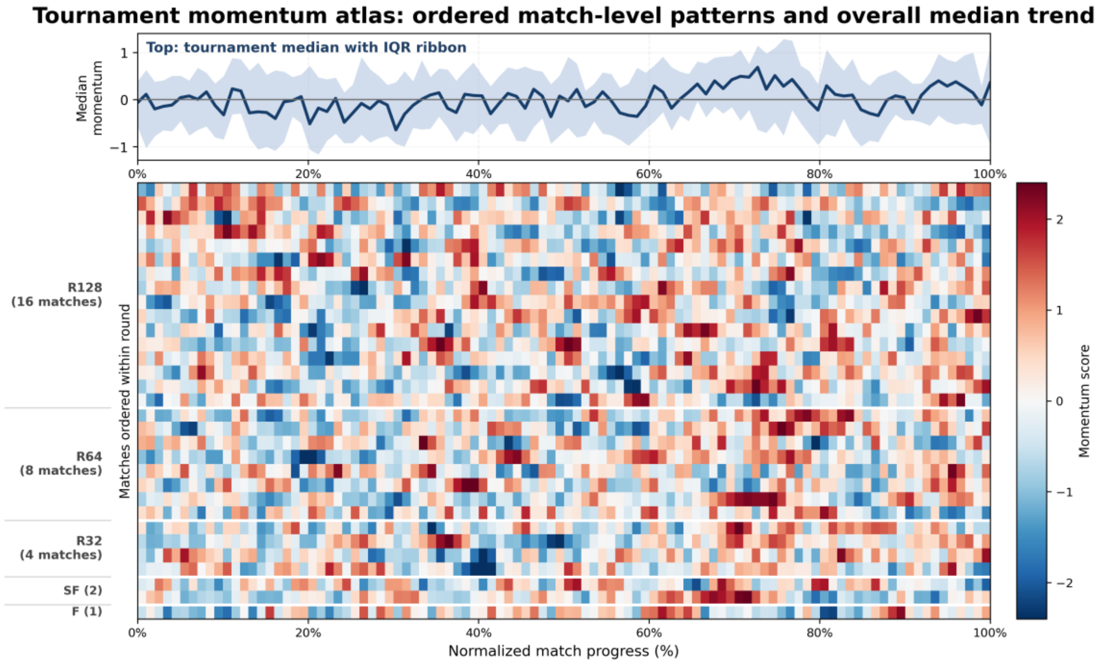

Figure 4. Tournament momentum atlas (Top: overall median momentum across normalized match progress with an interquartile ribbon. Bottom: round-stratified heatmap of within-match standardized momentum, ordered within round by average absolute momentum. The color bar is placed in a separate axis to avoid overlapping the round labels).

Predictive value of momentum

The key validation question is whether momentum improves prediction of the next point once serving and score context are already known. Table 4 and Figure 5 show that prediction is moderate rather than near perfect. A score-and-serve logistic baseline achieved an AUC of 0.683 with 0.673 accuracy. Adding EWMA momentum increased AUC to 0.687 and slightly improved the Brier score. A full contextual model including lagged momentum and short streak variables reached AUC = 0.690. The modest gains suggest that momentum contributes incremental information beyond serve and score, but its effect is marginal rather than game-changing. These values are intentionally interpreted conservatively rather than as evidence of deterministic match forecasting.

The coefficient plot in Figure 5 confirms that serving status is the dominant predictor, which is fully consistent with the broader tennis forecasting literature. The last-three-points streak also has a positive coefficient, whereas the longer EWMA momentum term has a negative coefficient in the full model after controlling for streaks and score. This should not be treated as evidence that momentum is harmful. Instead, it indicates that a broad smoothed advantage partly overlaps with score structure and may contain mean-reversion once short-run streaks and serve are controlled. This reading also aligns with earlier cautions against overinterpreting momentum as a purely self-reinforcing force in tennis.

Table 4. Group cross-validation performance for next-point prediction.

Model	Mean AUC	AUC SD	Accuracy	Brier score
Score + serve	0.683	0.020	0.673	0.2198
Add EWMA momentum	0.687	0.020	0.673	0.2195
Add short streak	0.684	0.020	0.673	0.2198
Full context	0.690	0.023	0.673	0.2192

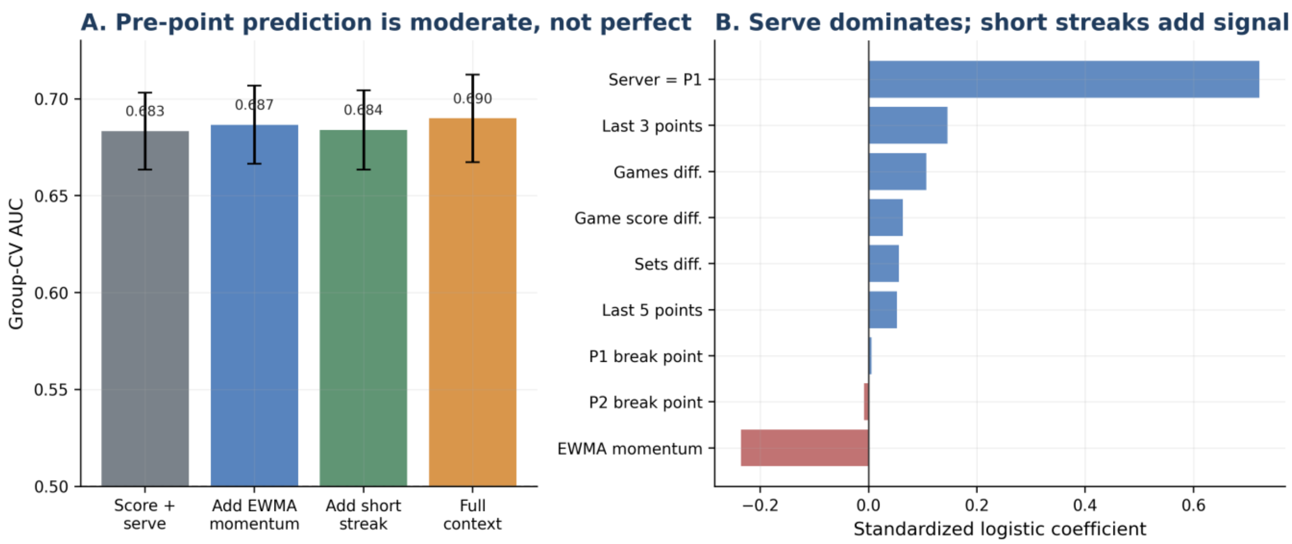

Figure 5. Conservative validation of next-point prediction (Panel A reports group-CV AUC; Panel B reports standardized coefficients from the full logistic model).

What creates large momentum swings?

Figure 6 decomposes the momentum score into its average absolute component contributions. Winning the point is the largest component by design, but the score is not simply a point-win counter. Winners and aces, errors avoided, break-point outcomes, net success, serve-return depth, speed, and movement balance all contribute additional information. This transparency is important because coaches can inspect each component instead of accepting a black-box score.

By isolating each contribution, coaches can pinpoint specific technical areas for targeted intervention, linking the metric directly to actionable training adjustments.

Figure 7 compares event rates on large swing points with all other points. Specifically, the data indicate that winners are about 60.2% of large swing points compared with 30.4% of other points. Aces occur on 25.2% of large swing points but only 6.2% of other points, and net-point wins are also enriched. This event profile helps explain why visible rallies often cluster after large one-step changes in operational momentum. Such clustering suggests that large momentum swings are accompanied by a higher frequency of decisive, high-risk shots, which in turn shapes the subsequent flow of the match.

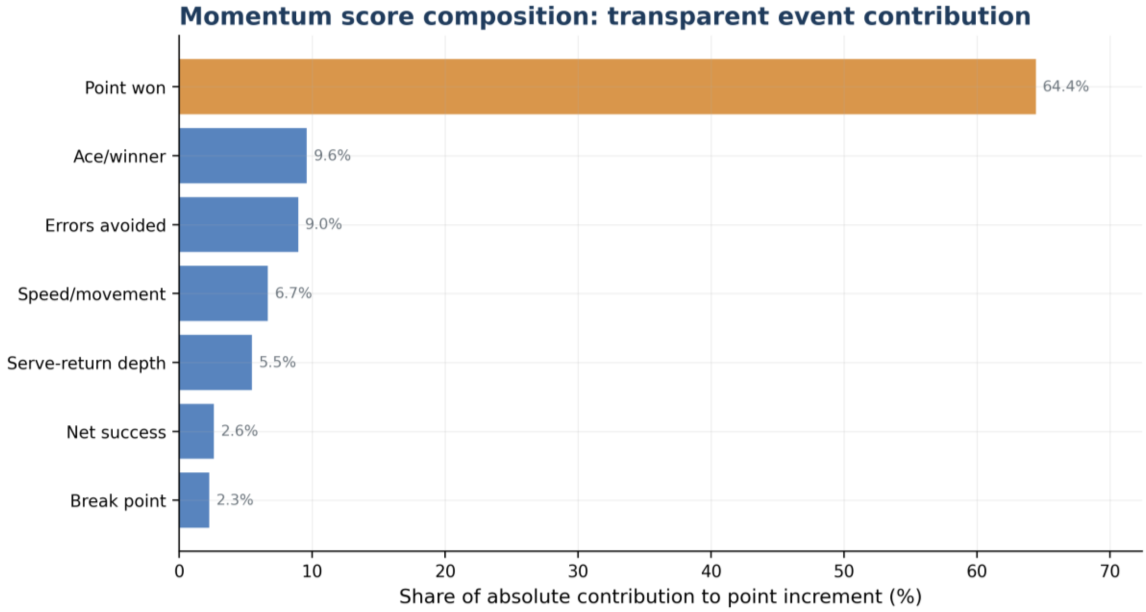

Figure 6. Composition of the point-performance increment (Bars show each component's share of the average absolute contribution).

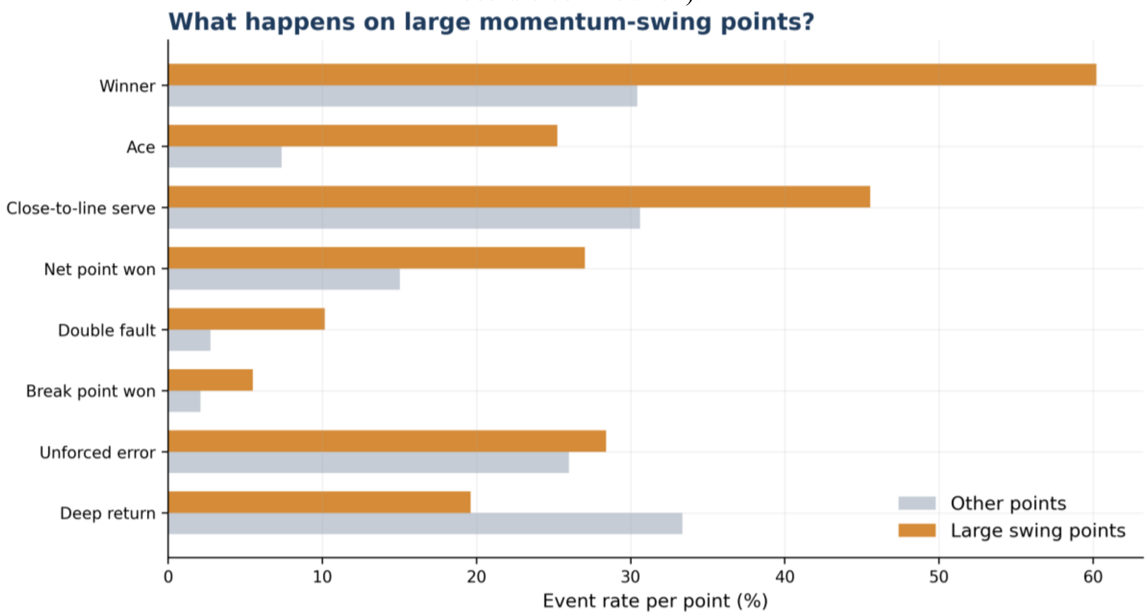

Figure 7. Event rates on large momentum-swing points compared with other points (Large swings are the top 10.0% of absolute one-step momentum changes).

Momentum state persistence

The state-transition matrix in Figure 8 shows that edge states are moderately persistent. From P1 edge, the next point remains P1 edge 61.2% and goes neutral 38.8%. From P2 edge, it stays P2 edge 59.4% and goes neutral 40.6%. Direct jumps between edge states are rare under ±0.65, as most reversals pass through neutral. This state-level view is more actionable than a single accuracy statistic, as it directly identifies when to press and when to regroup; coaches can use edge states to consolidate advantage and neutral states to shift momentum [10].

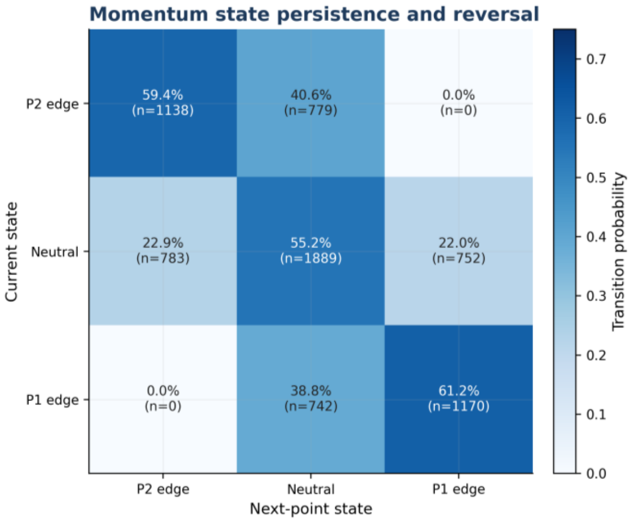

Figure 8. Momentum state transition probabilities (States are defined using standardized momentum thresholds of ±0.65).

Discussion

The revised framework changes the interpretation of the original paper in three ways. First, it replaces a task-based contest narrative with a journal-style research question and evidence chain. Second, it avoids treating momentum as a directly observed psychological entity. Third, it separates description from prediction. This separation matters. Much of the recent tennis-momentum literature can blur the boundary between descriptive pattern finding and genuine pre-point forecasting. This is especially true when complex machine-learning models are involved.

The results are consistent with the broader tennis analytics literature. Serve matters strongly, score context matters, and recent point sequences contain some information but not enough to justify deterministic claims of near-perfect predictability. They are also consistent with the view that momentum in tennis is partly strategic and partly perceptual: visible runs exist, but they coexist with match structure and can reverse quickly.

The operational momentum score remains useful despite its modest predictive gain. Its primary value is diagnostic. It highlights where a match changes direction, summarizes technical events into a single player-relative curve, and creates a common language for linking video review to data. This is especially helpful in applied settings where coaches need interpretable summaries rather than algorithmic scores with unclear meaning [11].

Practical implications for coaches

For coaching practice, the most useful output is not a claim that momentum guarantees future success. The useful output is a list of situations that should trigger attention. First, serve remains the main structural advantage. Momentum should therefore be reviewed in conjunction with serve status, not in isolation. A player who appears to be "on a run" while receiving may deserve special attention because that pattern is harder to sustain and therefore potentially more tactically informative.

Second, large swing points are enriched in winners, aces, close-to-line serves, net-point wins, double faults, and break-point conversions. Training plans should emphasize high-pressure serve routines, first-shot patterns after serve, and decision making near the net. These are the event types most associated with visible directional shifts. This interpretation is compatible with earlier evidence that perceived momentum in tennis often becomes salient around clustered high-leverage points rather than around average rallies.

Third, neutral momentum states are strategically valuable. In a neutral state, neither player has a strong recent edge, and a single high-quality point can move the match into an edge state. Players should be trained to recognize these moments and avoid low-percentage decisions that generate avoidable large swings.

Limitations

This study has several limitations. The dataset contains only Wimbledon 2023 Gentlemen's singles matches after the first two rounds. The results should not be generalized automatically to women's tennis, lower-level tournaments, different surfaces, or doubles. Weather, injury status, and player coaching interactions were not observed. The competition-style origin of the source data also means that some potentially useful contextual variables are absent even though the point-by-point record is rich.

The component weights are transparent but not universal. They are chosen to reflect tennis leverage and to produce a stable descriptive score. Future work could estimate weights from larger datasets using hierarchical models while retaining the leakage controls used here. The state thresholds also require sensitivity analysis in broader data. Finally, the working CSV was obtained through a public mirror of the COMAP dataset. Researchers should cite the original COMAP source and verify licensing or competition-use conditions before redistribution.

Conclusion

A publishable tennis momentum study should be transparent about what data are used, when variables become available, and what claims the model can support. Using 7,284 point records from 31 Wimbledon 2023 men's matches, this manuscript proposes an interpretable EWMA-based momentum score and evaluates its incremental predictive value under group cross-validation. The score is effective for visualizing match flow and identifying large swing points, but it adds only modest next-point predictive information beyond serve and score. This nuanced finding is more useful than an overfitted claim of near-perfect prediction.

The practical contribution is a reproducible framework that coaches and analysts can use to identify swing moments, review tactical events, and design training routines for pressure situations. The scientific contribution is a leakage-aware structure for studying momentum in tennis without confusing descriptive scoring with causal psychological inference.

Funding

This work was not supported by any funds.

Data availability

Reconstructed datasets and redrawn figures are documented in the supplementary tables and accompanying data package. The values were reconstructed from the source report using displayed equations, printed tables and embedded figure panels.

Acknowledgements

The author would like to show sincere thanks to those techniques who have contributed to this research.

Conflict of Interest

The author declares no conflict of interest.

References

[1] Meier, P., Flepp, R., Ruedisser, M. V., Franck, E. (2020) Separating psychological momentum from strategic momentum: evidence from men's professional tennis. *Journal of Economic Psychology*, 78, 102269.

[2] Lv, X., Gu, D., Liu, X., Dong, J., Li, Y. (2024) Momentum prediction models of tennis match based on CatBoost regression and random forest algorithms. *Scientific Reports*, 14, 18834.

[3] Wang, J., Guo, S., Zhou, Y. (2024) A multidimensional momentum chain model for tennis matches based on difference equations. *PLOS One*, 19(12), e0316542.

[4] Depken, C. A., Gandar, J. M., Shapiro, D. A. (2022) Set-level strategic and psychological momentum in best-of-three-set professional tennis matches. *Journal of Sports Economics*, 23(5), 598-623.

[5] Friligkos, G., Papaioannou, E., Kaklamanis, C. (2023) A framework for applying the Logistic Regression model to obtain predictive analytics for tennis matches. *Technium*, 15, 60.

[6] Xia, Y., Li, C., Zhang, T. (2025) Analyzing momentum shifts in tennis: a machine-learning approach to predicting match outcomes. *Applied Sciences*, 15(4), 2018.

[7] Liu, C., Yang, J., Cui, Y. (2025) Quantifying and predicting momentum in tennis match via machine learning approach. *International Journal of Racket Sports Science*, 7(1), 46-58.

[8] Honey, C. J., Mahabal, A., Bellana, B. (2023) Psychological momentum. *Current Directions in Psychological Science*, 32(4), 284-292.

[9] Zang, Q., Li, Z., Hu, J., Zhang, T., Zhang, W. (2024) Prediction of winning and losing in tennis match based on entropy weight-TOPSIS and machine learning model. *Transactions on Computer Science and Intelligent Systems Research*, 5, 1529-1535.

[10] Gupta, K., Krishnamurthy, V., Deb, S. (2025) What elements of the opening set influence the outcome of a tennis match? An in-depth analysis of Wimbledon data. *IIMB Management Review*, 37(1), 100519.

[11] Zhong, M., Liu, Z., Liu, P., Zhai, M. (2024) Searching for the effects of momentum in tennis and its applications. *Procedia Computer Science*, 242, 192-199.