"Exploring the impact of break points and contextual variables on point-by-point outcomes in ATP tennis matches"

Isabella Brown
MSc Sport Data Analytics
2024/25

CONTENTS

Abstract

Objective: Research shows pressure has both negative and positive effects on performance in sport. However, previous research on pressure in tennis ignores external influences that may amplify pressure effects. The purpose of this study was to examine how pressure situations, specifically break points, affect point outcomes in ATP tennis matches. It also investigated whether external factors such as court surface, tournament round and player ranking predict performance under pressure. The main hypothesis was that break points would be associated with an increased likelihood of server errors. **Methods:** Over 178,000 points were analysed across 53 ATP tournaments from 2017. Generalised linear mixed models (GLMMs) and random forests (RFs) were implemented to test both the effect of break points (pressure) on point outcomes and whether these points could be predicted with the variables court surface, tournament round, and player ranking differences. **Results:** GLMM results revealed that break points significantly increased the odds of the server losing the point (OR = 1.16, $p<.001$). Grass courts and hard courts reduced this risk (OR = 0.89 and OR = 0.96, respectively; both $p<.001$). The Round of 128 increased the odds of the server losing the point (OR = 1.11, $p<.05$) as did the Round of 64 (OR = 1.09, $p<.05$). No other predictors showed any significant effects. All GLMMs and RF models showed limited predictive performance (AUC ≈ 0.52). **Conclusions:** Break points significantly increase the chance of the server losing the point, highlighting the importance of psychological resilience under pressure. Although court surface and tournament round influenced general point outcomes, they, along with ranking differences, did not predict performance on break points. The limited predictive power of current models suggests a need for further research to develop more comprehensive models that incorporate additional psychological or contextual variables.

Keywords: Pressure, Attentional Control Theory (ACT), Generalised linear mixed model (GLMM), Random forest (RF), Tennis

Introduction

Throughout their sporting careers, elite athletes deal with circumstances that put strain on their capacity to cope [1]. Compared to athletes competing at lower levels, elite athletes usually face additional pressures [2]. Pressure refers to excessive or demanding expectations, either perceived or actual, placed on an individual to think or behave in a specific way. This experience often results in cognitive and emotional distress [3]. For many years, research in sport has focused on the connection between psychological stress and athletic performance since it has considerable importance for comprehending success and failure in competitive settings [4]. For elite athletes, the capacity to handle stress and perform well under pressure can be deemed crucial [5]. Tennis, with its dynamic interplay of physical skill and psychological resilience, presents a unique setting for understanding performance under pressure.

Performance Under Pressure

Numerous real-world scenarios involve psychological pressures that influence outcomes. For example, time constraints in professional settings can generate significant pressure [6], such as when a single interview is used to assess a candidate's employability, or when doctors must perform emergency procedures under urgent conditions [7]. Pressure in these circumstances ultimately leads people to perform worse than usual [7].

While psychological pressure affects decision-making across various domains, its impact is particularly pronounced in sport where the outcome of a single moment can determine success or failure. Pressure in sport involves the existence of situational rewards for optimal or exceptional performance, frequently coupled with substantial consequences for inadequate performance [4]. These performance-contingent rewards in addition to ego relevance, audience observation, and competitiveness, raise the significance of performing well, in turn, raising the pressure [4].

Furthermore, psychological pressure is closely linked to anxiety, a state that disrupts attention and cognitive processing, reducing the athlete's ability to focus on task-relevant cues [8], [9]. When athletes respond negatively to perceived pressure and underperform, it is considered to be "choking" [10], [11]. Research over the past two decades has explored why some sportsmen seem to be more prone to choking than others [12]. The most prominent explanations have been focused on distraction and self-focus models of choking [13], [14]. The underlying premise of these models is that optimal performance is disrupted by a changed cognitive process in the presence of anxiety [12]. Specifically, distraction theories suggest a focus on cues that are unrelated to the task at hand, thus lowering the capacity for attentional control [15], [16]. Qualitative research has shown support for distraction models; athletes who were under pressure expressed negative thoughts and attributed poor performance to such distracting factors [17]. Whereas self-focus theories suggest that concentrating on the sequential rules governing motor movements can lead to poorer performance [14], [18], [19]. This notion is supported by experimental investigations showing that individuals tend to choke when they focus on a motor task step-by-step [19]. These studies require participants to focus on the individual components of specific movements involved in the task; this heightened self-awareness disrupted their automatic execution of the task, resulting in a significant decline in performance [19]. Psychological pressure, linked to anxiety, can disrupt attention and cognitive processing, with distraction and self-focus models offering prominent explanations for why some athletes are more prone to choking under pressure.

The Attentional Control Theory (ACT) [20] offers a more comprehensive framework for understanding the cognitive mechanisms underlying anxiety and performance. This makes it more versatile in explaining different choking scenarios [21]. According to ACT, anxiety disrupts the balance between two attentional systems, the goal-directed system, responsible for focusing on task-relevant cues, and the stimulus-driven system, which responds to external or threat-related distractions [13]. This imbalance leads to a heightened focus on threat-related cues at the expense of task efficiency, impairing performance. While ACT primarily addresses trait anxiety, the Attentional Control Theory: Sport (ACTS) [22], extends these principles to dynamic competitive environments. ACTS introduces a bidirectional relationship between pressure and performance; it suggests that anxiety results from ongoing appraisals of the perceived cost and likelihood of failure, both shaped by prior performance feedback and situational pressure [23]. When these factors are elevated, they interact to increase anxiety, impair attentional control and reduce performance as attention shifts from task-related cues to threat-related ones [22], [24].

The ACTS provides a theoretical foundation for understanding the cognitive mechanisms underlying performance under pressure. For example, Basanovic et al. [25] revealed that as state anxiety increased, individuals who experienced a smaller decline in their ability to control attention exhibited a greater tendency to focus on negative information. This highlights the role of inhibitory control in anxiety-related cognitive processing. Empirical studies have since sought to test these theoretical propositions, investigating how anxiety and psychological factors manifest in competitive environments. Research reveals that high-pressure situations can both enhance [26-32] and impair [17], [19], [33], [34] performance depending on various psychological and contextual factors. Geukes et al. [33] discovered that only when under intense pressure, state anxiety and fear of failure were significantly linked to performance in basketball competitions, highlighting how high-pressure scenarios can exacerbate negative psychological effects. Similarly, and more recently, Mesagno et al. [34] reported that cognitive and somatic anxiety increased from low- to high-pressure, leading to decreased performance among Australian football players, with "chokers" displaying significant performance declines and irrational beliefs. However, these studies rely on selective, small, and convenient samples. Smaller samples may result in underpowered models for detecting effect sizes and potential type II errors, where the null hypothesis may not be rejected when it is false [35]. Additionally, convenience sampling introduces estimate bias due to poor generalisability, as easily accessible participants may not accurately represent the target population [36]. Nevertheless, the ACTS framework offers insight into how anxiety affects cognitive control and performance under pressure, with research showing negative performance outcomes depending on psychological and contextual factors.

On the other hand, some athletes experience 'clutch', which is the term for successful performance under crucial, pressured conditions [26]. Several well-known acclaimed sporting events illustrate clutch performance, such as in 2012, Manchester City won their first Premier League title thanks to an injury-time strike by Sergio Aguero, showcasing his extraordinary composure in a high-stakes moment [27]. Similarly, Michael Jordan's goal to win the 1998 National Basketball Association (NBA) title with under ten seconds remaining is frequently regarded as the epitome of clutch performance [28]. In tennis, Roger Federer's comeback from two sets down in the 2009 French Open against Tommy Haas, including saving a crucial break point, exemplifies resilience and poise in critical circumstances, ultimately leading to his career Grand Slam [29]. These instances not only demonstrate individual talent but also the mental toughness needed to perform at the top level. Within academic literature, Mihalyi et al. [32] analysed NBA shooting and play-by-play data, revealing that under pressure, top scorers transitioned into elite facilitators. This finding highlights the concept of clutch in decision-making, demonstrating how elite players adapt their roles in high-stakes situations to optimise team performance, though team dynamics mean shared accountability which lowers anxiety levels [37]. So, it is important to examine the gap in effects within individual tasks, which this study aims to do given that tennis is inherently an individual sport.

The difference between excelling under pressure and struggling may come down to an individual's mindset. Positive emotions are strong predictors of better work performance [38]. This is supported by evidence from a throwing accuracy task, where participants who were led to believe they were well-suited to perform under pressure, demonstrated significantly better performance in high-pressure conditions compared to low-pressure conditions. In contrast, the control group, who did not receive this suggestion showed no significant difference in performance across pressure conditions [30]. By helping to eliminate any other possible reasons for the results, the use of a control group enhances the study's clarity and validity [39]. Therefore, this study effectively highlights the influence of psychological priming in performance under stress; it supports the idea that confidence can mitigate the negative effects of pressure. Laborde et al. [40] upholds these findings in a tennis context, where twenty-eight tennis players performed a series of serves, separated by a pressure manipulation. They revealed that cortisol, along with self-confidence, were predictive of performance under pressure. However, the way pressure was induced was artificial, resulting in an inadequate reflection of the complexities of psychological phenomena in a real-world competitive scenario [41]. This research illustrates how confidence and psychological priming can boost performance under pressure, though experimental settings may oversimplify real-world dynamics.

The idea that some individuals perform better under pressure extends beyond controlled experiments to real-world competitive settings. Similarly, and more specific to the present study, Goff et al. [31] found that under increased match pressure, both male and female tennis players are more likely to hold serve. However, despite this evidence of clutch performance, the study's use of random effects does not account for unobserved factors that vary across games, such as wind, sunshine, or fatigue. If these factors are correlated with the main situational variable, they could distort the results [31], [42]. For example, an opponent's fatigue may lead to errors, making it seem like the server is performing well under pressure when, in fact, the outcome is influenced by the opponent's mistakes. Nonetheless, the study mitigates this concern by performing a placebo check when the set score is tied. The results show that when the set is tied, the probability of holding serve does not significantly change compared to other service games, confirming that serving for the set changes competitive incentives regardless of fatigue [31]. By accounting for this potential confounding factor, this study enhances the validity of the findings [43], supporting the notion that some athletes excel under pressure.

Contrary to the above findings, Cohen-Zada et al. [44] compared performance across 2010 Grand Slams in both low- and high-pressure scenarios and revealed male tennis players consistently choke under pressure. Their use of match fixed effects, which control for constant match conditions, such as weather, temperature, and surface, helps reduce estimation bias [45]. More contemporary evidence reinforces these conclusions, including Harris et al. [10], who analysed variations in pressure point-by-point during Grand Slams and demonstrated that the chance of a performance error increased significantly with pressure increases. The use of linear mixed effects models ensures effective handling of hierarchical data, such as points within games, and games within matches [46]. The operationalisation of key variables ensures objective measures, in turn, enhancing the reliability of their findings [47]. Both studies benefit from large Grand Slam datasets, which enhances the validity of their findings, ensuring results are based on real-world observations rather than theoretical assumptions. However, it is essential that when observing the effects of pressure in tennis, authors conceptualise pressure, otherwise they risk threats to the reliability and validity of their findings.

Conceptualising Pressure in Tennis

Tennis is one of the most popular racquet sports, played against pairs (doubles) or between two players (singles). It has a distinctive score system that progresses from 0 (love) to 15, 30, 40, and finally game point; a player needs to win by a minimum of two points. Serving alternates between players with each game. Matches are divided into sets, typically best-of-three or best-of-five, and players must win six games (by at least two) to earn a set. Surfaces like grass, clay and hard courts influence both the ball's bounce and pace.

Tennis is commonly known as the "mental game", a term that assumes a player's mindset as essential to their performance [48]. In tennis, pressure situations often arise due to specific match conditions that heighten the importance of performance [49] and amplify the potential cost of failure [50]. Certain moments in tennis disproportionately impact the probability of victory. A critical example being when a player serves to close out a set; this scenario consists of a set score of 5-4 or 6-5, where the leading player is serving, and it holds tremendous weight [31]. Successfully holding serve secures the set, whereas being broken returns the set to 'on serve', effectively nullifying the immediate advantage.

However, pressure in tennis is not constant, despite assumptions made when comparing low- and high-pressure conditions [44]. Cohen-Zada et al. [44] define game pressure by categorising a match as high-stakes if both players have previously won at least four games; otherwise, it is considered low-stakes. However, this classification overlooks dynamic in-game fluctuations, such as break points, which can significantly impact pressure levels. Harris et al. [10] effectively addressed this by ranking pressure from 1 to 5, 5 denoting the highest pressure. They defined high-pressure situations as points occurring near the conclusion of games, sets, and matches, as these moments had a greater direct impact on the overall outcome. Additionally, they decided that game points and break points amplified pressure, given their greater significance compared to earlier points in a game. Another study that successfully incorporated in-game fluctuations was conducted by Pokharel and Zhu [51], who identified a substantial relationship between low threat levels and unforced errors, indicating a conscious attempt to avoid errors under high-threat conditions and increased carelessness in low-threat situations. Threat was defined by proximity to losing the game, ranked from 0 (e.g., 15-0) to 3 (e.g., 15-40). However, their methodology did not account for external factors beyond the match. It is essential to recognise that previous literature has primarily examined the effects of pressure on performance, often neglecting contextual factors that could amplify these pressure moments. The current study aims to address this gap by considering variables such as player ranking, court surface, and tournament round, highlighting how key moments in tennis demand both mental and physical resilience.

Effect of Tournament Round

In some cases, it may be the significance of the actual match such as a Grand Slam final, rather than a single point that drives pressure. A notable example is Jana Novotna's collapse in the 1993 Wimbledon final, where she led 4-1 in the final set but lost to Steffi Graf, highlighting how even dominant players can choke under the weight of a high-stakes final [52]. Alongside the heightened opportunity for personal achievement, several factors may contribute to an athlete choking under pressure in a final, most notably, the allure of prize money [53], [54] and the intensified audience scrutiny [55-57].

A study on football players found that a $10 monetary incentive led to significantly faster 50-yard dash times compared to a $3 incentive, indicating that higher financial rewards can enhance performance [53]. However, these amounts are minimal compared to the substantial prize money offered in ATP tournaments. Whilst the immense financial stakes in professional tennis may serve as motivation, they may introduce heightened pressure. More aligned with ATP prize money, Ariely et al. [54] investigated performance in situations where individuals could earn nearly half of their village's average annual consumer expenditure. They showed that poorer performance resulted from the largest financial incentive, aligning with the downfall of Novotna in the Wimbledon final. However, since this research was conducted in an economic context rather than a sports setting, further investigation is warranted to determine whether the same results apply to athletic performance.

Secondly, the impact of audience is prevalent in sport. The BBC's coverage of the men's 2023 Wimbledon final peaked at 11.3 million viewers [58], creating a high-pressure atmosphere. Research by Böheim et al. [55] found larger NBA crowds reduced free throw success, especially for weaker players. The presence of choking in front of an audience remains the case even when the audience is supportive [56]. This pattern extends across multiple sports, as seen in Major League Baseball (MLB) where Jane [57] obtained further proof that the size of the crowd has an adverse causal influence on each player's chance of batting well, though star batters often excel in such moments demonstrating clutch performance. As tournaments progress, the audience size increases at each stage [58], amplifying the pressure on players. Given the evidence linking performance to both prize money and crowd size, this study uniquely incorporates tournament round as a key variable when analysing pressure's effect on point outcomes. However, it is crucial to consider that various factors amplify the importance of playing in a final, with prize money and audience size being the most prominent.

Effect of ranking

In the NBA, teams under playoff pressure are expected to win [32]; this aligns with tennis, where higher-seeded players are expected to win when facing lower-seeded or unseeded opponents, raising questions about pressure and performance. Conde-Ripoll et al. [59] revealed that in padel somatic anxiety and self-confidence were higher before competition than training matches among players with higher rankings. For low-rank players, only self-confidence was greater before competition. This confirms Mihalyi et al.'s [32] findings and suggests higher-ranked players may experience more anxiety prior to competition due to performance expectations. However, anxiety was assessed 30-45 minutes before matches using self-report questionnaires, which neither captures last-minute changes closer to competition nor in-game fluctuations; additionally, self-reported measures are susceptible to response bias [60]. Although English was chosen for the questionnaires this was not the participants native language and may have affected comprehension and response accuracy. Finally, the sample used only included ten players, limiting the generalisability of the findings [36].

On the other hand, findings from MLB [57] revealed the opposite, star players experienced clutch performance in front of large audiences compared to weaker players who experienced detrimental effects. Therefore, it is crucial to investigate player ranking in this study to determine whether highly ranked tennis players exhibit a similar effect. Unlike Harris et al.'s [10] anonymised analysis, this study includes player rankings to address that gap. Similarly, Jetter and Walker [61], using data from over 100,000 tennis matches, found that during significant events (such as Grand Slams), higher-ranked players increased their winning percentage both overall and in critical sets. When the incentives were highest, players with higher rankings were able to deliver clutch performances. These mixed results warrant further investigation into how player ranking influences point outcomes under pressure. To the researcher's knowledge, this is the first study to investigate this relationship.

Effect of court surface

Court surface is another important but often overlooked factor in performance under pressure. Clay courts are considerably slower than hard courts, which has been linked with heightened physiological reactions and greater physical and mental fatigue due to increased endurance demands [62], [63]. However, Kilit et al.'s [63] studied recreational players and Reid et al.'s [62] focused on players in a pro program, so findings may not generalise to elite players who likely have the conditioning to manage clay's demands. Furthermore, these studies do not account for performance on grass courts, leaving a gap in understanding across all surface types.

Grass courts are the fastest court surface [31]; Sim and Choi [64] found that when servers are at a disadvantage on grass courts compared to clay courts, each point has greater influence on the game's outcome. Conversely, when servers hold an advantage on clay courts, the importance of each point increases accordingly [64]. They also reported that on grass courts, points played at 30-40 or 40-Adv are particularly crucial, as the server has a greater chance of recovering from a one-point disadvantage because grass courts tend to favour servers [31]. However, the Markov chain used in this study treats all states equally and does not account for potential performance differences under pressure. Although, supporting their findings, Goff et al. [31] revealed higher service hold probabilities in matches played on grass courts and lower hold probabilities on slower clay surfaces, which tend to negate fast-moving serves. Research from Doğan et al. [65] confirms these findings, revealing the average number of aces to be lowest on clay courts. They also found that unforced errors were significantly lower on grass courts, indicating that players experience less pressure on this surface. However, this only accounts for Grand Slam data in its analysis and may not fully represent trends across all professional tournaments. Court surface may influence performance under pressure, with slower clay courts increasing physical and mental fatigue, and faster grass courts tending to favour servers and reduce pressure. However, existing models are limited by participant level and may not fully capture performance variations under pressure or across different tournaments. The current study addresses these limitations by examining performance across all major court surfaces using data from a large selection of ATP competitions.

Research Gap and Subsequent Aims

Previous literature has only observed pressure and the outcome [10], [31], [44], [51], without accounting for external factors, resulting in an incomplete understanding of pressure in elite tennis. Despite widespread recognition of the importance of pressure points, systematic approaches to predicting pressure point outcomes remain underdeveloped in academic literature especially in individual sports like tennis, where clutch performance is under-researched compared to team contexts. Therefore, the present study provides a novel approach to analysing point-by-point data from singles ATP tennis matches. Not only does it analyse pressure points, such as break points, therefore effectively examining momentary fluctuations rather than blocked pressure situations [13], but it also incorporates court surface, player rankings, and tournament round in pressure performance to create a more comprehensive model. It uses point-by-point data to assess the influence of these factors and predict outcomes, using training and test data to determine whether these variables (court surface, ranking, tournament round), can forecast future results. This study seeks to answer the following research questions:

1. How does pressure during break points, affect a player's ability to win a point in ATP matches?
2. What factors play a role in the point scoring ability of pressure points.

Based on previous literature this study comes up with the following hypotheses:

1. There will be a greater likelihood of errors with the presence of pressure points such as break points.

2. Factors such as tournament round, court surface, and player ranking will predict the outcome of pressure points.

By investigating these research questions and testing these hypotheses, this study aims to deepen understanding of how psychological pressure influences tennis performance at a detailed level. The findings will offer valuable insights into the interplay between external factors, such as court surface and tournament round, and internal factors, like player ranking and the psychological weight of crucial points. Ultimately, this research could contribute to performance optimisation strategies for players and coaches, enhancing decision-making under pressure and refining training methodologies to improve competitive resilience. Additionally, the insights may benefit broadcasters by enriching match narratives and could inform predictive models in the sport betting industry.

Methods

Data

This study retrospectively analysed point-level data from professional men's tennis matches to explore gameplay dynamics, including momentum, pressure situations and server-return interactions [10], [64], [66]. The dataset was obtained from the publicly available repository curated by Jeff Sackmann on GitHub (https://github.com/) [67]. GitHub's robust online interface supports transparent code editing, documentation, and peer review, enhancing data credibility and reliability through error correction and consistent updates [68]. Jeff Sackmann, a respected expert in sport statistics is widely cited by professionals [69] making him a reliable source for tennis data due to his extensive experience in data curation and analytics contributions [70-72]. Finally, the validity of the dataset was confirmed by cross-referencing match outcomes with official records.

The dataset provided detailed, sequential point-by-point data from 2,147 completed ATP matches across 56 tournaments, held between January 1, 2017, and August 31, 2017. Matches that were not completed due to player retirements were excluded from the original data to maintain consistency and accuracy, as these entries often contain incomplete or invalid score data [67].

Tournament round information and surface type were collected separately from each tournament's database. Player rankings were obtained from the ATP ranking website [73], matched to each player's most recent monthly ranking relative to the match date.

Prior to analysis, the dataset underwent extensive preprocessing in R Studio (Version 2024.12.1+563), involving cleaning, filtering, and reformatting for analytical compatibility and to ensure data integrity and usability.

The variables in Table I were included for preparation and analysis.

TABLE I

VARIABLES AND DESCRIPTIONS

Variable	Description
pbp	Point-by-point outcome (i.e. S or R denoting server or returner)
game_number	Game number
set_number	Set number
points	Point winner
scoreline	Scoreline
point_winner	Point winner corrected
pbp_id	Match ID
W_L_shifted_up	Point winner corrected to reflect actual occurrence
break_points	Break points
Round	Tournament round
Rank_Diff	Ranking difference
Surface	Surface type
Server_Name	Server name
Rank_Diff_Scaled	Ranking difference scaled for modelling

Data Cleaning and Preparation

Data cleaning guarantees high-quality data, improving both analysis and machine learning (ML) performance [74]. The initial phase of data cleaning involved identifying and addressing missing values as they limit the amount of information accessible to ML models during training, thus affecting classification accuracy [75]. The dataset contained no missing values. Davis Cup and Hopman Cup matches were excluded to focus on individual performance rather than team-based events. Additional variables, tournament round, surface type and player ranking were subsequently integrated into the original dataset.

The initial wide-format point-by-point data (i.e., 'SRSSRS;SSSRS…'), was transformed into a long format to facilitate analysis where each point was assigned its own row. A new 'scoreline' column was then created (i.e., 15-0, 15-15…). Matches with tiebreaks were identified by a "/" in the 'pbp' column and were excluded to simplify data processing due to their irregular scoring and rarity.

To track match progression, 'game_number' and 'set_number' columns were added. A dot "." in the 'points' column marked the end of a set, prompting a reset of the game number and an increment of the set number; changes in server names indicated new matches, resetting both counters.

To manage server alternation and scoring logic, data were split into odd- and even-numbered sets. A ‘point_winner’ column was created and adjusted based on game number parity – flipping winners in even games during odd sets and vice versa. The ‘scoreline’ column was updated accordingly.

After these corrections, the subsets were recombined and sorted by ‘pbp_id’ to restore match order. A new column ‘W_L_shifted_up’, was created to shift the outcomes upward by one row, aligning point results with their actual occurrence. These outcomes were recoded as binary values: 1 was assigned if the point was won by the returner, and 0 if it was won by the server.

Break points were used to identify pressure points because winning these points is more crucial than winning other points in match [10]. They were identified by the server being one point away from losing the game, (i.e., when the score was “X-40” or “40-Ad”). A value of 1 was assigned for break points, and 0 for all other points.

A final inspection revealed missing ‘point_winner’ values only where sets ended (marked by a “.”) and corresponding ‘W_L_shifted_up’ values due to the use of the lead() function. Missing game scores only occurred mid-game. Prior to analysis, rows with missing outcomes (‘W_L_shifted_up’) were removed to ensure integrity of statistical testing [75].

Data Analysis

Assumptions and Descriptive Statistics

The dataset’s structure was reviewed, and necessary adjustments were made to ensure that variables were appropriately formatted for analysis. Outliers were addressed early, as they can bias statistical estimates [76]. Boxplots were used to detect outliers in key variables. Outliers in player rankings were mitigated by grouping rankings into categories, since lower-ranked players tend to show greater variability and sparser data, which may distort model performance. The category “200+” was deemed acceptable as these rankings represented 7.50% of the data. Outlier values in ‘game_number’ and ‘set_number’ were linked to longer matches, particularly during Grand Slam tournaments. Furthermore, the ‘break_points’ variable demonstrated slight class imbalance, with break points (coded “1”) accounting for only 9.32% of the data.

Descriptive statistics were used to summarise the distribution of point-level outcomes across key predictors including court surface, tournament round, and player ranking.

Relationship between Outcome and Predictors

To examine the relationships between the outcome variable and key predictors, chi-square tests and correlation analysis were conducted. Chi-square tests of independence were performed to assess whether point outcomes (‘W_L_shifted_up’) were significantly associated with break points (‘break_points’), court surface (‘Surface’), and tournament round (‘Round’). These tests evaluated whether the observed distribution of point outcomes significantly differed from the expected distribution across categories of each predictor variable.

For the continuous predictor ranking difference (‘Rank_Diff’), a Pearson correlation was conducted to assess the strength and direction of its linear relationship with point outcomes, which were numerically encoded for this analysis.

Generalised Linear Mixed Model (GLMM)

This study employs GLMMs with a logistic link to analyse point-level outcomes in professional tennis. The hierarchical nature of tennis data, where points are nested within games, matches and individual players, violates the independence assumption of standard regression techniques [77]. GLMMs accommodate this structure by estimating fixed effects of interest while modelling random intercepts [78]. The binary outcome variable (point won or lost) necessitates logistic regression [79], while fixed effects are used to directly test the hypotheses: (1) pressure points reduce server win probability, and (2) these effects are moderated by player rankings, court surface, and tournament round. GLMMs outperform standard logistic regression, by accounting for data dependencies and reducing Type I error risk [80]. Their utility is supported by recent applications in sport analytics; for example, Blything and Blything [78] analysed the influence of timeouts in tennis, while Herold and Breuer [81] examined optimal sponsor placement in broadcasts. By modelling both fixed and random effects, GLMMs yield statistically robust estimates while maximising ecological validity. This is a critical advantage for deriving actionable insights from high-pressure competitive scenarios.

Before fitting a GLMM, the outcome variable was converted into a binary factor indicating whether the server won or lost the point. The continuous predictor ‘Rank_Diff’ was standardised using z-scaling to address scaling sensitivity in logistic regression [82]. The dataset was split into training (80%) and testing (20%) subsets, and a fixed random seed (123) was used to guarantee reproducibility of the split.

Model 1: Pressure only predictor

A mixed-effects logistic regression model was fitted on the training data, using 'break_points' as the main predictor with random intercepts for 'pbp_id' and 'Server_Name' to account for contextual variability (e.g., court conditions) and individual heterogeneity (e.g., performance under pressure). Model performance was assessed using coefficients, odds ratios with 95% confidence intervals (CI), and predicted probabilities on the test set. Discriminative ability was assessed using the receiver operating characteristic (ROC) curve and the area under the curve (AUC). The optimal probability threshold was selected using Youden's index [83]. Predictions were converted to binary labels and compared to actual outcomes using a confusion matrix, from which precision, recall, and F1 score were derived.

Model 2: Full predictor set

The second model extended the first by including additional predictors: surface, round and scaled ranking difference. The same evaluation metrics and procedures as Model 1 were applied to assess performance on the test data.

Model 3: Interaction effects

This model tested whether the effects of break points on point outcome was moderated by surface, round, or ranking difference through interaction terms. As before, performance was evaluated using ROC curves, AUC and classification metrics. Additionally, predicted marginal effects of break points across surface types were visualised.

Random Forest (RF)

Complementing the GLMM approach, RFs were chosen to overcome the limitations of parametric models in capturing the complex dynamics of pressure points in tennis. RFs are well-suited to this context due to their ability to capture non-linear relationships and higher-order interactions without the manual specification required by GLMMs [84]. They also handle correlated features [85] and offer feature importance metrics [86]. Their ensemble structure, aggregating multiple decision trees, makes them resistant to overfitting, thereby yielding more precise and reliable insights from hierarchical tennis data structure [87], [88]. Additionally, the operational simplicity and computational efficiency of this method make it highly regarded [89]. Prior research has shown RFs can achieve over 70% accuracy when forecasting tennis match outcomes [90-93].

Model 4: Pressure only RF

A baseline RF model was trained using 'break_points' as the sole predictor of point outcome. The model was configured to grow 500 decision trees with one variable considered at each split. Test set prediction were evaluated using a confusion matrix, precision, recall, F1 score, ROC curve and AUC.

Model 5: All predictors RF

A second RF model was developed using all available predictors: ‘break_points’, ‘Surface’, ‘Round’ and ‘Rank_Diff_scaled’. As before, 500 trees were grown, and variable importance was computed. The model was evaluated using the same procedures as Model 4.

Finally, a comparative table summarised model performance across all approaches, reporting AUC, accuracy, sensitivity and specificity.

Follow-Up Analysis

To further explore model performance under high-pressure scenarios, a subset of the data containing only break points was created. This allowed focused examination of player performance in critical match outcomes.

Again, the ‘Rank_Diff’ variable was scaled to ensure comparability across predictors. The filtered dataset was then partitioned into training and test sets using an 80/20 split with a fixed random seed to ensure reproducibility.

GLMM: Full model on break points

A GLMM was fitted with ‘Surface’, ‘Round’, and ‘Rank_Diff_scaled’ as fixed effects, and random intercepts for ‘pbp_id’ and ‘Server_Name’. Model performance was evaluated on the test set. Predicted probabilities were generated and used to construct an ROC curve and calculate the AUC. A confusion matrix was generated to assess classification performance, with a Youden’s index selected threshold.

RF: Full model on break points

An RF was also trained on the same filtered dataset, including all predictors. The model was trained using 500 trees with default settings for node splitting and feature sampling. Performance was assessed on the test set using predicted probabilities, ROC curve, AUC and a confusion matrix.

Data analysis was completed using R Studio (Version 2024.12.1+563). Statistical significance was accepted when $p<0.05$.

Results

Descriptive Statistics

The clean dataset comprised of 178,443 points across 1,252 matches, 268 different players and 53 tournaments. Player rankings ranged from 1 to 1953. The majority of matches were played on hard courts (49.28%, see Table II).

TABLE II.

MATCHES PLAYED BY COURT SURFACE

Surface	No of Matches	Proportion (%)
Clay	477	38.10
Grass	158	12.62
Hard	617	49.28

Matches were distributed across various tournament rounds, with the highest proportion occurring in the Round of 32 (34.19%) followed by the Round of 16 (18.77%). Table III provides a full breakdown of match distribution by round. For a more detailed breakdown of matches by surface and round see Table AI in Appendix A.

TABLE III.

MATCHES PLAYED BY ROUND

Round	No of Matches	Proportion (%)
Final	33	2.64
Semi Finals	62	4.95
Quarter Finals	129	10.30
Round of 16	235	18.77
Round of 32	428	34.19
Round of 64	221	17.65
Round of 128	144	11.50

Players were grouped into ranking categories. The largest share of matches involved players ranked 51-100 (30.08%), followed by those ranked 11-30 (19.28%). Full details are in Table IV.

TABLE IV.

MATCHES PLAYED BY RANKING CATEGORY

Ranking Category	No of Matches	Proportion (%)
Top 10	237	10.58
11–30	432	19.28
31–50	418	18.65
51–100	674	30.08
101-200	312	13.92
200+	168	7.50

Break Point Win Rates

Figure 1 presents a heatmap visualising break point win rate across surfaces and rounds. Warmer colours indicate higher break point conversion percentages. Full details on break point win rates, broken down individually by surface, round, and server ranking category are provided in Appendix A (Tables AII-AIV).

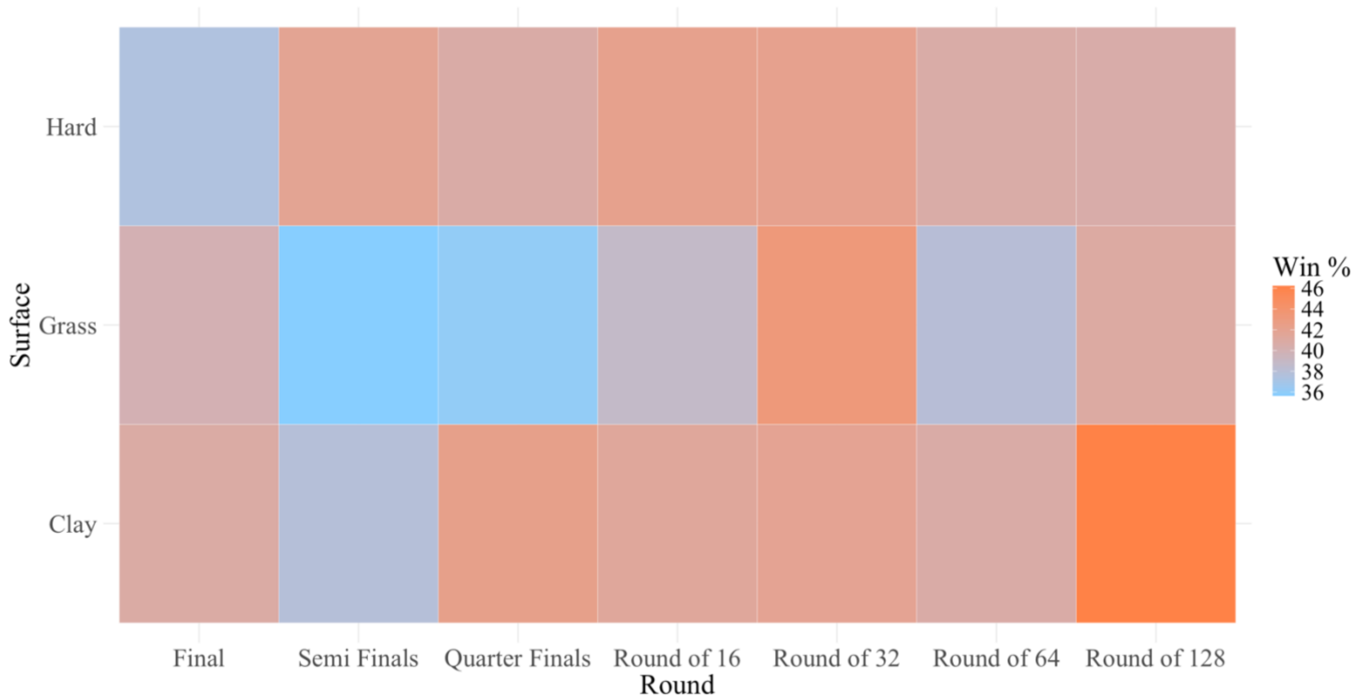

Fig. 1. Heatmap of break point win rates across surface and round

Relationship between outcome and predictors

The results (Table V) suggested that break points were strongly associated with point outcomes. Additionally, there was a statistically significant relationship between point outcome and court surface. While tournament round also demonstrated a significant association with point outcomes, the effect was comparatively weaker. In contrast, the difference in player rankings exhibited no significant correlation with the win/loss outcome.

TABLE V.

RELATIONSHIP BETWEEN OUTCOME AND PREDICTORS

Variable Compared	Test	Test Statistic	df	p
W_L_shifted_up vs break_points	Chi-squared	$X^2 = 97.03$	1	< 0.001***
W_L_shifted_up vs Surface	Chi-squared	$X^2 = 62.04$	2	< 0.001***
W_L_shifted_up vs Round	Chi-squared	$X^2 = 16.50$	6	0.01*
W_L_shifted_up vs Rank_Diff	Pearson correlation	t = -0.04	178440	0.97

*(Notes: ***p<.001 **p<.01 *p<.05)*

Generalised Linear Mixed Model (GLMM)

Model 1: Pressure only predictor

The fixed effect of 'break_points' was found to be positive and highly significant (p<.001) with an estimate of .15 and odds ratio (OR) = 1.16, with 95% CI [1.13, 1.20]. This implies that when a point is a break point, the odds of the server losing the point increases by 16% compared to non-break points. There was a small but meaningful variance at both the match level ('pbb_id', SD = 0.08) and the server level ('Server_Name', SD = 0.12), justifying the inclusion of random intercepts to account for heterogeneity across players and matches.

Model performance on the test set showed weak practical predictive power. The confusion matrix (Figure B1 in Appendix B) achieved 48.63% accuracy, with a 95% CI [48.11%, 49.15%]. The model's sensitivity 38.93% was and specificity was 64.57%; the precision was 64.34% and the F1 score was 48.51%. The AUC was 0.52, only slightly better than random guessing (AUC = 0.5). Cohen's Kappa was 0.03 indicating no level of agreement between prediction and the actual outcome [94].

Model 2: Full predictor set

To improve performance and account for additional contextual variables, a second GLMM was estimated using multiple predictors. The model converged successfully using the 'bobyqa' optimiser. Random effects estimates showed modest variability attributable to 'pbp_id' ($\sigma^2 = 0.004$, SD = 0.06) and 'Server_Name' ($\sigma^2 = 0.013$, SD = 0.11). Table VI summarises the fixed effects estimates, OR, and 95% CI.

TABLE VI.

GLMM FULL PREDICTOR SET RESULTS

Predictor	Estimate	OR	95% CI for OR	p
(Intercept)	-0.55	0.58	[0.50, 0.65]	< 0.001***
Break Points	0.15	1.16	[1.13, 1.20]	< 0.001***
Grass Courts	-0.12	0.89	[0.85, 0.93]	< 0.001***
Hard Courts	-0.05	0.96	[0.92, 0.98]	< 0.001***
Quarter Finals	0.05	1.05	[0.97, 1.33]	0.23
Round of 128	0.10	1.11	[1.03, 1.19]	0.012*
Round of 16	0.05	1.05	[0.97, 1.13]	0.24
Round of 32	0.07	1.08	[1.00, 1.15]	0.06
Round of 64	0.09	1.09	[1.02, 1.17]	0.02*
Semi Finals	0.04	1.04	[0.95, 1.13]	0.41
Rank Difference	0.001	1.00	[0.99, 1.01]	0.85

*(Notes: ***p<.001 **p<.01 *p<.05)*

Break points significantly increased the odds of the server losing the point by 16%. Both grass and hard courts reduced the likelihood of the server losing compared to clay courts, with grass showing a stronger protective effect. Early rounds such as the Round of 128 and Round of 64 were associated with higher odds of the server losing the point, though effect sizes were modest; later rounds did not reach statistical significance. Rank difference was not a significant predictor.

Model performance was evaluated on a test dataset. Even though this model showed slightly improved performance (AIC = 189069.0 compared to Model 1's AIC = 189097.9), the AUC was 0.52, again indicating predictions barely stronger than chance. The confusion matrix (Figure B2 in Appendix B) showed 52.63% accuracy with a 95% CI [52.11%, 53.15%] and κ=0.03, along with recall of 55.57% and specificity of 47.82%. The precision was 63.62% and the F1 score was 59.32%. Overall, the model's predictive performance was limited.

Model 3: Interaction effects

The third model extended Model 2 by incorporating interaction terms between 'break_points' and other predictors. The main effects were consistent with model 2; grass courts significantly reduced the odds of the server losing compared to clay by 11% (OR = 0.89, p<.001), as did hard courts by 5% (OR = 0.95, p<.001). Early tournament rounds (Round of 128, Round of 64) increased the odds of the sever losing (OR = 1.10, OR = 1.09, respectively; p<.05) while later rounds were non-significant. Ranking differences remained non-significant.

The interaction effects for break points x surface were non-significant. Similarly, the interaction effects for break points x round and break points x ranking difference were non-significant. See Figure C1 in Appendix C for exploration of interactions between break points and surface.

The model showed worse performance than Model 2 (AIC = 189078.2 vs. 189069.0). The AUC was 0.52, reinforcing limited predictive gains. The confusion matrix (Figure B3 in Appendix B) displayed sensitivity of 55.18% , specificity of 47.86% and accuracy of 52.41% with a 95% CI [51.89%, 52.93%] and κ=0.03. the F1 score was 59.04% and the precision was 63.47%. Model performance was not enhanced by these interaction effects.

Random Forest (RF)

Model 4: Pressure only RF

This RF model was trained solely on break points. This resulted in an AUC of 0.5, which is equivalent to random guessing. The confusion matrix (Figure B4 in Appendix B) revealed accuracy of 62.15% with 95% CI [61.65%, 62.66%] and κ=0. The precision was 62.2% and the F1 score was 76.7%, however, this model failed to generalise, achieving 100% recall but 0% specificity by predicting only the majority class (Server Wins).

Model 5: All predictors RF

A second RF model was trained using all predictors to assess whether the inclusion of multiple contextual factors would enhance the predictive performance. This model still showed no predictive power with an AUC of 0.53 and κ=0.0002. Precision and the F1 score were identical to the first model (62.2% and 76.7%, respectively). The accuracy was 62.16% with 95% CI [61.65%, 62.66%]. However, 0.03% specificity was achieved, with 99.99% recall.

Comparing performance of Model 1-5

The performance metrics of the five models are summarised in Table VII. The GLMM based models (1-3) showed balanced but modest performance. RF models (4-5) demonstrated overfitting to the majority class.

TABLE VII.

COMPARISON OF MODEL (1-5) PERFORMANCE

Model	AUC	Accuracy	Sensitivity	Specificity
Model 1: GLMM Basic	0.52	0.49	0.39	0.65
Model 2: GLMM All Predictors	0.52	0.53	0.56	0.48
Model 3: GLMM Interactions	0.52	0.52	0.55	0.48
Model 4: RF Basic	0.50	0.62	1.00	0.00
Model 5: RF All Predictors	0.53	0.62	1.00	0.00

Follow-Up Analysis

GLMM: Full model on break points

A GLMM was fitted to examine whether court surface, tournament round, or ranking differences predicted point outcomes during break points (N = 16,794). Estimates of random effects revealed that 'pbp_id' ($\sigma^2 = 0.004$, SD = 0.07) and 'Server_Name' ($\sigma^2 = 0.013$, SD = 0.12) were responsible for a moderate amount of variability. None of the predictors significantly affected the probability of the server losing the point on break point.

Predictive performance on the test set was also poor. The AUC was 0.51 and the optimal threshold for converting predicted probabilities to class labels was 0.43. The confusion matrix (Figure B6 in Appendix B) showed an accuracy of 56.67% with 95% CI [54.97%, 58.36%] and κ=0.03. Model sensitivity was 82.88% and specificity was 19.51%.

RF: Full model on break points

An RF classification model was also trained on the break point subset; however, it demonstrated limited practical predictive value. As shown in the confusion matrix (Figure B7 in Appendix B) the model achieved an overall accuracy of 58.64% with 95% CI [56.95%, 60.31%]. Importantly, κ=0, indicating no improvement over chance-level classification. While the model attained sensitivity of 100%, its specificity was 0%, misclassifying all instances of the minority class. The AUC was 0.51, suggesting the model's ability to distinguish between the two outcome classes was no better than random guessing.

Discussion

According to previous literature, this is the first study to analyse how pressure points (break points) and external factors (tournament round, court surface and player rankings) affect tennis performance. It aimed to increase knowledge of the specific ways in which psychological stress impacts tennis players. The hypotheses were (1) that there will be a greater likelihood of errors with the presence of pressure points and (2) that the external factors above will predict the outcome of pressure points. The principal finding was that break points significantly increased the odds of the server losing the point; this finding was consistent across all models suggesting pressure negatively affects server performance, leading to the acceptance of hypothesis 1. While surface and round influenced outcomes, rankings did not, and predictive models performed poorly, leading to rejection of hypothesis 2.

More specifically, chi-square tests confirmed significant associations between break points, surface and round with point outcome; however, their inability to model the hierarchical structure of tennis data limits their interpretability [95]. This limitation was addressed through a GLMM which provided not only statistical associations but also meaningful effect sizes. Crucially, the GLMM results underscore the robust and systematic influence of break points on performance, showing a 16% increase in the likelihood of the server losing the point under pressure. This finding empirically supports psychological theories of performance under pressure, such as distraction, self-focus and ACTs [13], [14], [20], suggesting that break points function as psychological stressors that impair motor execution and decision making. It also aligns with previous research demonstrating that high-stakes moments affect cognitive and emotional regulation, potentially impairing athletic performance [10], [25], [33], [34], [44]. These insights have important practical implications for player development and coaching, emphasising the need to develop psychological resilience and effective coping strategies in high-stakes scenarios. Techniques such as pre-shot routines and positive self-talk draw on the ACTS framework to reduce negative evaluations of mistakes and alleviate situational anxiety [23].

Court surface was significantly associated with point outcomes, with grass and hard courts offering a slight protective effect for servers compared to clay. This reflects biomechanical factors, as faster ball speeds and lower bounces on these surfaces favour aggressive serving and reduce break opportunities [31]. Goff et al. [31] similarly reported higher service hold probabilities on grass, likely due to reduced returner reaction time and increased ace potential. However, this relationship is neither static nor universal. Returners adopting aggressive court positions can reduce serve dominance on grass and hard courts [96], challenging the notion that surface type alone governs match dynamics. Furthermore, court conditions vary with environmental factors and maintenance practices across tournaments [97], adding further variability to surface-dependent outcomes. Future research should explore how moderating variables interact with surface type, rather than treating it as a fixed determinant of outcomes.

Similarly, tournament round also exhibited a statistically significant association with point outcomes, with servers more likely to lose points in earlier rounds (Rounds of 128 and 64). This contradicts prior literature suggesting that performance deteriorates under high-stakes conditions involving greater financial incentives [54] or larger audiences [55], [57], typically found in later rounds. This discrepancy likely arises because Ariely et al. [54] investigated performance in economic tasks, which do not generalise to high-skill competitive sport. Unlike those facing a single high-reward scenario, professional tennis players regularly compete for monetary gains, potentially normalising this pressure. Another explanation is that later rounds predominantly feature higher-ranked players who may be less susceptible to pressure, as lower-ranked players are eliminated earlier. This aligns with Jane's [57] observation that elite batters performed better under larger crowds. Early rounds may include more unseeded or less experienced players, possibly disrupting expected serving advantages. Finally, reduced perceived pressure in early rounds may result in under-arousal in line with the Yerkes-Dodson law [98], which suggests that both low and excessive arousal can impair performance. These findings suggest a non-linear relationship between stress and athletic performance. Consequently, coaches should implement training to avoid complacency in early rounds and regulate arousal to the optimal mid-range, as outlined by the Yerkes-Dodson framework.

Contrary to expectations, player ranking differences were not significantly associated with point outcomes. This was surprising given previous Grand Slam tennis research showing higher-ranked players typically have higher match-winning percentages [61]. A likely explanation for this discrepancy lies in the level of analysis. While Jetter and Walker [61] examined match-level outcomes, the present study focused on individual points, suggesting rankings may better predict overall match success than isolated point outcomes. Furthermore, the inclusion of random effects for players in the GLMM may have accounted for a substantial portion of variance attributable to individual skill, thereby diminishing the observable effect of rankings. Another methodological consideration concerns the operationalisation of ranking, where in this study individual player rankings were collapsed into a single ranking-difference metric. This approach does not distinguish between scenarios such as a match between players ranked 70 and 100 versus those ranked 1 and 31, cases that may vary greatly in competitiveness. As a result, the analysis was also unable to test whether 'star' players exhibit clutch performance under pressure, a phenomenon observed by Jane [57] in MLB. Future research could investigate non-linear modelling approaches such as neural networks (NN), to better capture subtle performance dynamics [99], particularly among top-ranked players where small differences may reflect significant disparities in ability. Artificial NNs have already shown promise in tennis betting strategies [100]. Additionally, examining whether rankings are more predictive at intermediate levels of granularity, such as games or sets, could help identify thresholds where ranking becomes a more reliable performance indicator.

Notably, interaction effects between break points and other variables were non-significant indicating that the influence of break points on point outcomes remained consistent regardless of surface type, tournament round, or ranking differences. This suggests that break points exert uniform psychological pressure, irrespective of external conditions. This pattern is congruent with ACT, along with distraction and self-focus models, which propose that high pressure moments impair task-focused processing regardless of context. Also, under pressure, individuals may attempt to consciously control automated skills, further disrupting performance [9]. These findings warrant further investigation into internal influences such as, individual differences like resilience (e.g., measured using the Ego-Resiliency Scale, which assesses the ability to adapt to emotional and situational demands and demonstrates acceptable reliability (α=.76) [101]), or player experience (e.g., match exposure, career longevity), as potential moderators not addressed in the present analysis.

Despite the presence of individually significant predictors in the GLMM, both the GLMM and RF models demonstrated limited predictive power, with AUC values only marginally above chance (AUC$\approx$0.52). This underscores a fundamental distinction between explanatory power (statistical significance) [102] and predictive performance. Although break points, surface, and tournament round are associated with point outcomes, these factors did not translate into reliable predictions for new data. Additionally, when the analysis focused solely on break points, none of the predictors (surface, round, ranking difference) significantly influenced the probability of the server losing the point. A plausible explanation for this is that tennis point outcomes are influenced by a complex array of situational, psychological, and performance-related factors not captured by the available variables; these unknowns add noise, reducing the model's ability to generalise [103]. For example, psychological pressure is not exclusive to break points and may occur at any moment, varying by individual in ways that cannot be inferred from match metadata alone.

Emerging literature supports this interpretation. Physiological studies have demonstrated that match-induced pressure reduces heart rate variability (HRV) [104] and increases heart rate and brainwave activity [105], [106], indicating elevated cognitive and emotional arousal. Fuentes-García et al. [104] measured HRV 24 hours prior to matches and again 20 minutes before match play. However, this approach does not capture in-game physiological fluctuations reflecting moment-to-moment changes in pressure. While Pineda-Hernández [106] used imagery techniques to simulate varying levels of match pressure, these scenarios remain artificial and may not fully replicate the complexity of real-life match conditions, where pressure is influenced by a combination of dynamic, contextual and interpersonal factors. Nevertheless, these findings provide evidence that pressure experiences are not limited to break points. Future research should consider in-game HRV monitoring as a practical and non-invasive method to assess real-time stress responses. Models could also benefit from incorporating players' self-reported perceptions of pressure as physiological markers alone do not fully capture the subjective experience [107]. However, such data is difficult to gather due to privacy concerns and the sensitive nature of psychological metrics [108], and even when obtained, their individualised nature complicates generalisable prediction.

Furthermore, pressure may change with momentum [32]. For example, research has shown that bathroom timeouts, often taken after losing a set, can significantly alter match momentum [78], creating a new form of psychological pressure as players attempt to maintain or regain control [109]. The strategic use of timeouts adds a layer of mental complexity, reflecting how pressure can emerge from the dynamics of match flow, not just isolated high-stakes points. This has important coaching implications; instead of attempting to predict outcomes, coaches should emphasise resilience in break point scenarios, adaptability across surfaces and psychological readiness across tournament rounds. For example, mental tools such as counteracting thoughts, routines, and breathing techniques have been found to regulate heart rate and brain activity to optimal levels [106], supporting performance under varying forms of pressure.

In addition to psychological and contextual factors that are not captured in the available data, structural aspects of tennis also limit predictive performance, even with accurate data and well-specified models. A key issue is the inherent imbalance in point outcomes, where servers tend to win the majority of points, particularly on faster surfaces like grass or hard courts [110]. This creates class imbalance, where one outcome (server winning) dominates, posing challenges for classification algorithms known to struggle with such distributions [111]. Although this reflects the reality of professional tennis, it results in models that appear adequate by predicting the majority class but lack sensitivity to minority cases, partly explaining the low AUC values observed. This study prioritised ecological validity by preserving the natural distribution of point outcomes, avoiding techniques like random oversampling (ROS) that artificially rebalance data by duplicating minority classes. While this choice reflects real-world match conditions, it limits the model's ability to address class imbalance directly. To address this and improve sensitivity, future models could incorporate techniques like ROS, which has performed well in other domains, such as heart disease classification, with reported accuracy rates of 0.8-0.9 and recall between 0.69-0.8 [112].

Strengths

This study offers several notable strengths enhancing its rigour and relevance. First, it uses a large, representative dataset comprising over 178,000 individual points from more than 1,000 matches, addressing limitations of earlier research based on much smaller datasets [33], [34]. Additionally, the inclusion of data from 53 different ATP tournaments, rather than only Grand Slams as in prior studies [10], [44], [61], [65], increases both statistical power [113] and the generalisability of the findings across various competitive contexts.

Second, by analysing real match data this study maintains high ecological validity by capturing authentic pressure scenarios in professional tennis. Ecological validity is crucial to sport performance research, as it ensures that findings reflect the genuine cognitive and behavioural responses elicited in competitive environments [114]. Also, the data is sourced from a publicly available repository, which ensures transparency [115]; detailed data cleaning and preprocessing steps are documented, facilitating reproducibility.

Third, this study takes a novel, multidimensional approach by incorporating court surface, tournament round, and player ranking, offering a more comprehensive understanding of performance under pressure. Unlike previous research that treated pressure more statically [10], [31], [40], [44], this context-sensitive analysis better reflects performance dynamics. Additionally, grounding the study in ACT and its sport-specific extension (ACTS) alongside distraction and self-focus models, provides a strong theoretical foundation within a well-established cognitive model that reveals the complex interplay between anxiety, attention, and performance in elite settings [13], [14], [20], [22-24].

Methodologically, the application of GLMMs accounts for tennis's nested and hierarchical data structure and controls for individual- and match-level variation [78]. Additionally, the inclusion of RF models adds a machine learning perspective capable of detecting complex, non-linear patterns [89]. Although predictive power was limited, the findings offer practical value for players and coaches, highlighting the need for targeted pressure management strategies, particularly in break point situations and early tournament rounds.

Implications

This study advances theoretical understanding of performance under pressure in individual sports like tennis. By demonstrating that break points significantly increase the likelihood of server errors, it supports key elements of ACT, ACTS, distraction and self-focus models [13], [14], [23], reinforcing the view that pressure impairs performance by disrupting attentional control. Importantly, this study extends these frameworks by showing that the effect of pressure is consistent across surface types, tournament rounds and ranking differences, suggesting that pressure exerts a universally disruptive influence rather than being contingent on these specific contextual variables. This adds depth to existing theories, indicating that internal cognitive responses to pressure may be more influential than external match conditions.

Beyond theory, this study provides meaningful practical insights for athletes, coaches, and sport psychologists. Even though the models could not predict point outcomes, the consistent negative impact of break point pressure suggests that players are more vulnerable in these moments, warranting focused intervention. Coaches can use these findings to design training programs that simulate high-pressure scenarios, particularly break points, to help athletes develop coping mechanisms such as routines, and controlled breathing [106], [116]. Additionally, employing emotional regulation strategies, such as cognitive reappraisal, can enable individuals to reinterpret heightened arousal as performance-enhancing rather than threatening [117], [118].

The results also suggest that court surface and tournament round influence general point outcomes, indicating that tactical preparation should be adapted to these conditions. For example, coaches should focus on improving first-serve accuracy to protect against weaker second serves on fast surfaces, while returners need mental training to maintain composure and develop tactical adaptability when break chances are rare. Also, to prevent early round complacency, coaches should help players manage arousal levels, aiming for an optimal mid-range as suggested by the Yerkes-Dodson law. For example, players could learn to shift attention from 'winning easily' to executing specific techniques such as hitting 70% of first serves to corners. Moreover, if players are under-aroused, they could implement pre-match activation drills such as explosive jumps and sprints to elevate heart rate; if they are over-aroused, they should use controlled breathing techniques to stabilise focus [116].

Finally, while this study provides meaningful insights into how pressure affects performance, its limited predictive power reduces its utility for commercial applications such as betting or live broadcasting. For betting applications, where precision and reliability are essential, these findings highlight the limitations of relying solely on surface-level match metadata. Similarly, for broadcasters aiming to build predictive narratives during live coverage, the inability to accurately forecast point outcomes underscores the need to frame pressure as a compelling performance factor rather than a deterministic predictor. While this study informs performance strategy, it does not support point-level forecasting for commercial and entertainment use.

Limitations and Future Directions

Despite its strengths, this study has several limitations that should be acknowledged. First, although break points are widely regarded as high-pressure moments in tennis [10], [51], they are an imperfect measurement for psychological pressure. Players may experience pressure in other contexts such as match points, tiebreaks and momentum shifts, which were not captured here. This highlights another limitation; the analysis excluded tiebreaks to simplify data processing, potentially omitting some of the most psychologically intense phases of a match [119]. Also, individual differences in perceived pressure, shaped by experience, personality, or match context, were not considered. Future research could adopt mixed methods, integrating match data with qualitative insights from player interviews as well as physiological data such as heart rate measures [104-106] and cortisol [40] to better capture the internal experience of pressure. Real time monitoring of physiological responses could offer a more dynamic and individualised measure of pressure.

Furthermore, the study tested GLMMs and RFs, but these models demonstrated limited predictive power. This may reflect the absence of important psychological factors such as confidence, fatigue or real-time stress responses that are known to influence outcomes beyond the variables included in this study [25], [33], [34], [40]. Future models should therefore consider a broader range of variables including fatigue indicators, physiological responses and tactical data (e.g., serve placement, shot selection). This would enhance validity by accounting for more variance in point outcomes under pressure, thereby improving predictive performance with richer models. Further to this, alternative approaches such as Bayesian hierarchical models or NNs could more effectively capture non-linear performance dynamics [99], [120]. Future work should compare these methods to optimise predictive performance.

Another limitation of this study is that potential confounding variables such as weather conditions, crowd behaviour, player fatigue, injury status and opponent quality were not available in the dataset and thus could not be controlled for. These variables may have distorted the present findings [121]. For instance, a server's decline in performance during break points may be partly due to their fatigue at the end of service games. While random effects in the GLMM controlled for some unobserved heterogeneity, missing variables likely reduced explanatory power. Incorporating head-to-head records or in-match fatigue indicators, as well as qualitative information about the weather, crowd and injury status could mitigate these confounds and enhance model validity.

Importantly, this study focused exclusively on ATP (men's) matches, limiting the generalisability of the findings to WTA (women's) tournaments. Given well-documented disparities in prize money, playing styles and psychological stressors between genders [44], it remains uncertain whether similar pressure responses occur in women's matches. Future research incorporating WTA tournaments as well could help determine whether similar or differing pressure responses occur, therefore yielding richer insights.

Finally, Grand Slam matches (best-of-five sets) were not analysed separately, potentially obscuring unique pressure effects specific to major tournaments, where stakes (reputation and prize money), match duration and psychological demands differ substantially from regular tour events [10]. Additionally, the dataset was limited to matches from 2017, raising concerns about temporal relevance. Since January 2025, the introduction of revised on-court and off-court coaching means that coaches can issue advice at any time including change of ends and set breaks [122]; this includes brief access to player analysis technology [123]. These regulatory changes could influence player behaviour and decision-making under pressure, with coaches being able to remind players of coping strategies during a match. Therefore, the applicability of the present findings to current match contexts is limited and future studies should account for this evolving context.

Conclusion

In conclusion, the current study provides novel insights into performance under pressure in tennis, specifically during break points. Overall, there was a significant increase in the odds of the server losing when it was a break point, supporting theories of pressure-induced performance disruption. Interestingly, court surface, tournament round and player ranking differences did not influence break point outcomes, although court surface and tournament round did affect general point outcomes; grass and hard courts increased server success while early tournament rounds hindered server success. However, none of these factors improved the predictive performance of the GLMMs and RF models tested, both of which demonstrated poor predictive accuracy. These results underscore the complexity of tennis performance and the need for more comprehensive models that incorporate a broader range of variables. It would be more relevant for future studies to consider analysis of matches from January 2025 onwards since the implementation of new coaching rules that may alter player behaviour under pressure. Additionally, the findings emphasise the importance of cognitive resilience for players serving under pressure, especially during break points. Ultimately, this study contributes to the theoretical understanding of pressure in sport and offers valuable insights for coaches, sport psychologists and researchers alike.

References

[1] E. McLoughlin, D. Fletcher, G. M. Slavich, R. Arnold, and L. J. Moore, ‘Cumulative lifetime stress exposure, depression, anxiety, and well-being in elite athletes: A mixed-method study’, *Psychol. Sport Exerc.*, vol. 52, p. 101823, Jan. 2021, doi: 10.1016/j.psychsport.2020.101823.

[2] R. Arnold, D. Fletcher, and K. Daniels, ‘Demographic differences in sport performers’ experiences of organizational stressors’, *Scand. J. Med. Sci. Sports*, vol. 26, no. 3, pp. 348–358, Mar. 2016, doi: 10.1111/sms.12439.

[3] American Psychological Association, ‘Pressure’. Accessed: Mar. 21, 2025. [Online]. Available: https://dictionary.apa.org/pressure

[4] R. F. Baumeister and C. J. Showers, ‘A review of paradoxical performance effects: Choking under pressure in sports and mental tests’, *Eur. J. Soc. Psychol.*, vol. 16, no. 4, pp. 361–383, Oct. 1986, doi: 10.1002/ejsp.2420160405.

[5] F. C. Klingner, B. C. H. Huijgen, and M. Kempe, ‘The heat is on: Investigating the effect of psychological pressure on competitive performance in elite surfing’, *J. Sports Sci.*, vol. 41, no. 6, pp. 596–604, Mar. 2023, doi: 10.1080/02640414.2023.2230032.

[6] Q. Cao, J. Li, and X. Niu, ‘Tempus fugit: The impact of time constraint on investor behavior’, *J. Econ. Behav. Organ.*, vol. 200, pp. 67–81, Aug. 2022, doi: 10.1016/j.jebo.2022.05.022.

[7] A. Bucciol and A. Castagnetti, ‘Choking under pressure in archery’, *J. Behav. Exp. Econ.*, vol. 89, p. 101581, Dec. 2020, doi: 10.1016/j.socec.2020.101581.

[8] A. Nieuwenhuys and R. R. D. Oudejans, ‘Anxiety and perceptual-motor performance: toward an integrated model of concepts, mechanisms, and processes’, *Psychol. Res.*, vol. 76, no. 6, pp. 747–759, Nov. 2012, doi: 10.1007/s00426-011-0384-x.

[9] L. J. Roberts, M. S. Jackson, and I. H. Grundy, ‘Choking under pressure: Illuminating the role of distraction and self-focus’, *Int. Rev. Sport Exerc. Psychol.*, vol. 12, no. 1, pp. 49–69, Jan. 2019, doi: 10.1080/1750984X.2017.1374432.

[10] D. J. Harris, S. J. Vine, M. W. Eysenck, and M. R. Wilson, ‘Psychological pressure and compounded errors during elite-level tennis’, *Psychol. Sport Exerc.*, vol. 56, p. 101987, Sep. 2021, doi: 10.1016/j.psychsport.2021.101987.

[11] C. Mesagno, D. Marchant, and T. Morris, ‘Alleviating Choking: The Sounds of Distraction’, *J. Appl. Sport Psychol.*, vol. 21, no. 2, pp. 131–147, May 2009, doi: 10.1080/10413200902795091.

[12] C. Mesagno, J. Garvey, S. J. Tibbert, and P. Gröpel, ‘An Investigation Into Handedness and Choking Under Pressure in Sport’, *Res. Q. Exerc. Sport*, vol. 90, no. 2, pp. 217–226, Apr. 2019, doi: 10.1080/02701367.2019.1588935.

[13] D. J. Harris *et al.*, ‘The effect of performance pressure and error-feedback on anxiety and performance in an interceptive task’, *Front. Psychol.*, vol. 14, p. 1182269, May 2023, doi: 10.3389/fpsyg.2023.1182269.

[14] R. Sullivan, L. Uiga, R. S. W. Masters, G. Anson, and A. Nieuwenhuys, ‘Conscious motor processing and the pressure-performance relationship: a systematic review’, *Int. Rev. Sport Exerc. Psychol.*, vol. 17, no. 2, pp. 1079–1104, Dec. 2024, doi: 10.1080/1750984X.2022.2127327.

[15] T. G. Lee and S. T. Grafton, ‘Out of control: Diminished prefrontal activity coincides with impaired motor performance due to choking under pressure’, *NeuroImage*, vol. 105, pp. 145–155, Jan. 2015, doi: 10.1016/j.neuroimage.2014.10.058.

[16] Z. Liu and R. Tang, ‘Choking under pressure: positive cues can mitigate choking in behavior inhibition and switching’, *Curr. Psychol.*, vol. 43, no. 48, pp. 36700–36713, Dec. 2024, doi: 10.1007/s12144-024-07098-8.

[17] D. M. Hill and G. Shaw, ‘A qualitative examination of choking under pressure in team sport’, *Psychol. Sport Exerc.*, vol. 14, no. 1, pp. 103–110, Jan. 2013, doi: 10.1016/j.psychsport.2012.07.008.

[18] L. Schücker, N. Hagemann, and B. Strauss, ‘Attentional Processes and Choking under Pressure’, *Percept. Mot. Skills*, vol. 116, no. 2, pp. 671–689, Apr. 2013, doi: 10.2466/30.25.PMS.116.2.671-689.

[19] P. Gröpel and C. Mesagno, ‘Choking interventions in sports: A systematic review’, *Int. Rev. Sport Exerc. Psychol.*, vol. 12, no. 1, pp. 176–201, Jan. 2019, doi: 10.1080/1750984X.2017.1408134.

[20] M. W. Eysenck, N. Derakshan, R. Santos, and M. G. Calvo, ‘Anxiety and cognitive performance: Attentional control theory.’, *Emotion*, vol. 7, no. 2, pp. 336–353, 2007, doi: 10.1037/1528-3542.7.2.336.

[21] N. B. Carlisle, ‘Flexibility in Attentional Control: Multiple Sources and Suppression’, *Yale J. Biol. Med.*, vol. 92, no. 1, pp. 103–113, Mar. 2019.

[22] M. Eysenck and M. Wilson, ‘Sporting performance, pressure and cognition. Introducing attentional control theory: sport’, in *An Introduction to Applied Cognitive Psychology*, 2nd ed., Psychology Press, 2016, pp. 330–351.

[23] D. J. Harris, S. J. Vine, M. W. Eysenck, and M. R. Wilson, ‘To err again is human: exploring a bidirectional relationship between pressure and performance failure feedback’, *Anxiety Stress Coping*, vol. 32, no. 6, pp. 670–678, Nov. 2019, doi: 10.1080/10615806.2019.1643459.

[24] A. J. Cocks, R. C. Jackson, D. T. Bishop, and A. M. Williams, ‘Anxiety, anticipation and contextual information: A test of attentional control theory’, *Cogn. Emot.*, vol. 30, no. 6, pp. 1037–1048, Aug. 2016, doi: 10.1080/02699931.2015.1044424.

[25] J. Basanovic, I. Kaiko, and C. MacLeod, ‘Change in Attentional Control Predicts Change in Attentional Bias to Negative Information in Response to Elevated State Anxiety’, *Cogn. Ther. Res.*, vol. 45, no. 1, pp. 111–122, Feb. 2021, doi: 10.1007/s10608-020-10176-3.

[26] M. J. Schweickle, S. A. Vella, and C. Swann, ‘Exploring the “clutch” in clutch performance: A qualitative investigation of the experience of pressure in successful performance’, *Psychol. Sport Exerc.*, vol. 54, p. 101889, May 2021, doi: 10.1016/j.psychsport.2021.101889.

[27] S. Hart, ‘Celebrate Sergio: Soccer’s most clutch players’, Soccer Guide. Accessed: Feb. 28, 2025. [Online]. Available: https://www.soccer.com/guide/celebrate-sergio-soccers-clutch-players

[28] E. Engler, ‘Top ten moments from Michael Jordan’s career’, Pipe Dream. Accessed: Feb. 28, 2025. [Online]. Available: https://www.bupipedream.com/ac/top-ten-moments-from-michael-jordans-career/134478/

[29] A. Sokolowski, ‘June 1, 2009: The day Roger Federer kept his dreams of Roland-Garros glory alive’, Tennis Majors. Accessed: Feb. 28, 2025. [Online]. Available: https://www.tennismajors.com/roland-garros-news/1st-june-2009-the-day-roger-federer-kept-his-dreams-of-roland-garros-glory-alive-261144.html

[30] B. McKay, R. Lewthwaite, and G. Wulf, ‘Enhanced Expectancies Improve Performance Under Pressure’, *Front. Psychol.*, vol. 3, 2012, doi: 10.3389/fpsyg.2012.00008.

[31] B. Goff, D. Wilson, and D. Zimmer, ‘Performing in high-pressure situations: the case of tennis players up a break and serving for the set’, *Appl. Econ.*, pp. 1–10, Sep. 2024, doi: 10.1080/00036846.2024.2381129.

[32] B. Mihalyi, G. Biczók, and L. Toka, ‘Momentum Matters: Investigating High-Pressure Situations in the NBA Through Scoring Probability’, in *Machine Learning and Data Mining for Sports Analytics*, vol. 2035, U. Brefeld, J. Davis, J. Van Haaren, and A. Zimmermann, Eds., in Communications in Computer and Information Science, vol.

2035. , Cham: Springer Nature Switzerland, 2024, pp. 77–90. doi: 10.1007/978-3-031-53833-9_7.

[33] K. Geukes, J. T. Harvey, A. Trezise, and C. Mesagno, ‘Personality and performance in real-world competitions: Testing trait activation of fear of negative evaluation, dispositional reinvestment, and athletic identity in the field’, *Psychol. Sport Exerc.*, vol. 30, pp. 101–109, May 2017, doi: 10.1016/j.psychsport.2017.02.008.

[34] C. Mesagno, S. J. Tibbert, E. Buchanan, J. T. Harvey, and M. J. Turner, ‘Irrational beliefs and choking under pressure: A preliminary investigation’, *J. Appl. Sport Psychol.*, vol. 33, no. 6, pp. 569–589, Nov. 2021, doi: 10.1080/10413200.2020.1737273.

[35] B. Nayak, ‘Understanding the relevance of sample size calculation’, *Indian J. Ophthalmol.*, vol. 58, no. 6, p. 469, 2010, doi: 10.4103/0301-4738.71673.

[36] J. Jager, D. L. Putnick, and M. H. Bornstein, ‘II. MORE THAN JUST CONVENIENT: THE SCIENTIFIC MERITS OF HOMOGENEOUS CONVENIENCE SAMPLES’, *Monogr. Soc. Res. Child Dev.*, vol. 82, no. 2, pp. 13–30, Jun. 2017, doi: 10.1111/mono.12296.

[37] S. A. Wolf and K. Utesch, ‘Everything is cool when you’re part of a team? The effects of outcome interdependence on appraisal, emotions, and performance under pressure’, *Psychol. Sport Exerc.*, vol. 74, p. 102683, Sep. 2024, doi: 10.1016/j.psychsport.2024.102683.

[38] L. I. Catalino, S. B. Algoe, and B. L. Fredrickson, ‘Prioritizing positivity: An effective approach to pursuing happiness?’, *Emotion*, vol. 14, no. 6, pp. 1155–1161, Dec. 2014, doi: 10.1037/a0038029.

[39] M. M. Pithon, ‘Importance of the control group in scientific research’, *Dent. Press J. Orthod.*, vol. 18, no. 6, pp. 13–14, Dec. 2013, doi: 10.1590/S2176-94512013000600003.

[40] S. Laborde, F. Lautenbach, M. S. Allen, C. Herbert, and S. Achtzehn, ‘The role of trait emotional intelligence in emotion regulation and performance under pressure’, *Personal. Individ. Differ.*, vol. 57, pp. 43–47, Jan. 2014, doi: 10.1016/j.paid.2013.09.013.

[41] G. A. Holleman, I. T. C. Hooge, C. Kemner, and R. S. Hessels, ‘The “Real-World Approach” and Its Problems: A Critique of the Term Ecological Validity’, *Front. Psychol.*, vol. 11, p. 721, Apr. 2020, doi: 10.3389/fpsyg.2020.00721.

[42] T. A. Gormley and D. A. Matsa, ‘Common Errors: How to (and Not to) Control for Unobserved Heterogeneity’, *Rev. Financ. Stud.*, vol. 27, no. 2, 2014.

[43] B. M. D'Onofrio, A. Sjölander, B. B. Lahey, P. Lichtenstein, and A. S. Öberg, 'Accounting for Confounding in Observational Studies', *Annu. Rev. Clin. Psychol.*, vol. 16, no. 1, pp. 25–48, May 2020, doi: 10.1146/annurev-clinpsy-032816-045030.

[44] D. Cohen-Zada, A. Krumer, M. Rosenboim, and O. M. Shapir, 'Choking under pressure and gender: Evidence from professional tennis', *J. Econ. Psychol.*, vol. 61, pp. 176–190, Aug. 2017, doi: 10.1016/j.joep.2017.04.005.

[45] M. Collischon and A. Eberl, 'Let's Talk About Fixed Effects: Let's Talk About All the Good Things and the Bad Things', *KZfSS Köln. Z. Für Soziol. Sozialpsychologie*, vol. 72, no. 2, pp. 289–299, Jun. 2020, doi: 10.1007/s11577-020-00699-8.

[46] R. H. Baayen, D. J. Davidson, and D. M. Bates, 'Mixed-effects modeling with crossed random effects for subjects and items', *J. Mem. Lang.*, vol. 59, no. 4, pp. 390–412, Nov. 2008, doi: 10.1016/j.jml.2007.12.005.

[47] C. Andrade, 'A Student's Guide to the Classification and Operationalization of Variables in the Conceptualization and Design of a Clinical Study: Part 1', *Indian J. Psychol. Med.*, vol. 43, no. 2, pp. 177–179, Mar. 2021, doi: 10.1177/0253717621994334.

[48] S. A. Kovalchik and M. Reid, 'Measuring clutch performance in tennis', *Stat. Appl.*, vol. 30, no. 2, pp. 255–268, Aug. 2018, doi: 10.26398/IJAS.0030-011.

[49] R. F. Baumeister, 'Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance.', *J. Pers. Soc. Psychol.*, vol. 46, no. 3, pp. 610–620, 1984, doi: 10.1037/0022-3514.46.3.610.

[50] H. Berenbaum, R. J. Thompson, and E. M. Pomerantz, 'The relation between worrying and concerns: The importance of perceived probability and cost', *Behav. Res. Ther.*, vol. 45, no. 2, pp. 301–311, Feb. 2007, doi: 10.1016/j.brat.2006.03.009.

[51] S. Pokharel and Y. Zhu, 'Analysis and Visualization of Sports Performance Anxiety in Tennis Matches', in *Advances in Visual Computing*, vol. 11241, G. Bebis, R. Boyle, B. Parvin, D. Koracin, M. Turek, S. Ramalingam, K. Xu, S. Lin, B. Alsallakh, J. Yang, E. Cuervo, and J. Ventura, Eds., in Lecture Notes in Computer Science, vol. 11241. , Cham: Springer International Publishing, 2018, pp. 407–419. doi: 10.1007/978-3-030-03801-4_36.

[52] H. Davis, 'Choking: the case of Jana Novotna: "One of the most famous collapses in sport"', Believe Perform. Accessed: Jan. 30, 2025. [Online]. Available: https://members.believeperform.com/choking-the-case-of-jana-novotna-one-of-the-most-famous-collapses-in-sport/#:~:text=collapses%20in%20sport%E2%80%9D-

,Choking%3A%20The%20Case%20of%20Jana%20Novotna%3A%20%E2%80%9COne%20of%20the,most%20famous%20collapses%20in%20sport%E2%80%9D&text=In%201993%2C%20Jana%20Novotna%20played,the%20third%20and%20final%20set.

[53] S. P. Jones, 'The Relationship between Monetary Incentives and Athletic Performance among Adolescent Males', *Walden Univ.*, May 2018, Accessed: Mar. 08, 2025. [Online]. Available: https://www.proquest.com/docview/2047544329?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

[54] D. Ariely, U. Gneezy, G. Loewenstein, and N. Mazar, 'Large Stakes and Big Mistakes', *Rev. Econ. Stud.*, vol. 76, no. 2, pp. 451–469, Apr. 2009, doi: 10.1111/j.1467-937X.2009.00534.x.

[55] R. Böheim, D. Grübl, and M. Lackner, 'Choking under pressure – Evidence of the causal effect of audience size on performance', *J. Econ. Behav. Organ.*, vol. 168, pp. 76–93, Dec. 2019, doi: 10.1016/j.jebo.2019.10.001.

[56] K. Harb-Wu and A. Krumer, 'Choking under pressure in front of a supportive audience: Evidence from professional biathlon', *J. Econ. Behav. Organ.*, vol. 166, pp. 246–262, Oct. 2019, doi: 10.1016/j.jebo.2019.09.001.

[57] W. Jane, 'Choking or excelling under pressure: Evidence of the causal effect of audience size on performance', *Bull. Econ. Res.*, vol. 74, no. 1, pp. 329–357, Jan. 2022, doi: 10.1111/boer.12307.

[58] 'Facts and Figures / FAQ'. Accessed: Mar. 08, 2025. [Online]. Available: https://www.wimbledon.com/en_gb/atoz/faq_and_facts_and_figures.html

[59] R. Conde-Ripoll, A. Escudero-Tena, and Á. Bustamante-Sánchez, 'Position and ranking influence in padel: somatic anxiety and self-confidence increase in competition for left-side and higher-ranked players when compared to pressure training', *Front. Psychol.*, vol. 15, p. 1393963, Apr. 2024, doi: 10.3389/fpsyg.2024.1393963.

[60] B. Lira *et al.*, 'Large studies reveal how reference bias limits policy applications of self-report measures', *Sci. Rep.*, vol. 12, no. 1, p. 19189, Nov. 2022, doi: 10.1038/s41598-022-23373-9.

[61] M. Jetter and J. K. Walker, 'Game, set, and match: Do women and men perform differently in competitive situations?', *J. Econ. Behav. Organ.*, vol. 119, pp. 96–108, Nov. 2015, doi: 10.1016/j.jebo.2015.07.017.

[62] M. M. Reid, R. Duffield, G. M. Minett, N. Sibte, A. P. Murphy, and J. Baker, 'Physiological, Perceptual, and Technical Responses to On-Court Tennis Training on

Hard and Clay Courts', *J. Strength Cond. Res.*, vol. 27, no. 6, pp. 1487–1495, Jun. 2013, doi: 10.1519/JSC.0b013e31826caedf.

[63] B. Kilit, E. Arslan, and Y. Soylu, 'The effect of court surfaces on physiological responses and game activities of recreational senior tennis players', *Pensar En Mov. Rev. Cienc. Ejerc. Salud*, vol. 22, no. 2, p. e58406, Jul. 2024, doi: 10.15517/pensarmov.v22i2.58406.

[64] M. K. Sim and D. G. Choi, 'The Winning Probability of a Game and the Importance of Points in Tennis Matches', *Res. Q. Exerc. Sport*, vol. 91, no. 3, pp. 361–372, Jul. 2020, doi: 10.1080/02701367.2019.1666203.

[65] İ. Doğan, S. Revan, and Ş. Arıkan, 'Analysis of Tennis Competitions on Different Court Surfaces', *Turk. J. Sport Exerc.*, vol. 23, no. 1, pp. 60–66, May 2021.

[66] Y. Cui and M. Ruano, 'Analysis of Point-by-Point Performance in Tennis: An Example of Novak Djokovic', in *11th World Congress of Performance Analysis of Sport*, Universidad de Alicante, 2017. doi: 10.14198/jhse.2017.12.Proc3.14.

[67] J. Sackmann, 'tennis_pointbypoint'. Sep. 14, 2017. Accessed: Oct. 10, 2024. [Online]. Available: https://github.com/JeffSackmann/tennis_pointbypoint/blob/master/pbp_matches_atp_main_current.csv

[68] P. H. P. Braga *et al.*, 'Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution', *Methods Ecol. Evol.*, vol. 14, no. 6, pp. 1364–1380, Jun. 2023, doi: 10.1111/2041-210X.14108.

[69] 'Jeff Sackmann', Speaker Profile. Accessed: May 09, 2025. [Online]. Available: https://www.sloansportsconference.com/people/jeff-sackmann

[70] S. A. Kovalchik, J. Sackmann, and M. Reid, 'Player, official or machine?: uses of the challenge system in professional tennis', *Int. J. Perform. Anal. Sport*, vol. 17, no. 6, pp. 961–969, Nov. 2017, doi: 10.1080/24748668.2017.1410340.

[71] R. M. Rodenberg, J. Sackmann, and C. Groer, 'Tennis integrity: a sports law analytics review', *Int. Sports Law J.*, vol. 16, no. 1–2, pp. 67–81, Jul. 2016, doi: 10.1007/s40318-016-0096-x.

[72] B. Wright, M. R. Rodenberg, and J. Sackmann, 'Incentives in Best of N Contests: Quasi-Simpson's Paradox in Tennis', *Int. J. Perform. Anal. Sport*, vol. 13, no. 3, pp. 790–802, Dec. 2013, doi: 10.1080/24748668.2013.11868689.

[73] ‘ATP Rankings’. Accessed: Mar. 11, 2025. [Online]. Available: https://www.atptour.com/en/rankings/singles?rankRange=0-5000&dateWeek=2017-01-30

[74] I. F. Ilyas and X. Chu, *Data Cleaning*. New York, NY, USA: ACM, 2019. doi: 10.1145/3310205.

[75] A. Palanivinayagam and R. Damaševičius, ‘Effective Handling of Missing Values in Datasets for Classification Using Machine Learning Methods’, *Information*, vol. 14, no. 2, p. 92, Feb. 2023, doi: 10.3390/info14020092.

[76] S. K. Kwak and J. H. Kim, ‘Statistical data preparation: management of missing values and outliers’, *Korean J. Anesthesiol.*, vol. 70, no. 4, p. 407, 2017, doi: 10.4097/kjae.2017.70.4.407.

[77] J. K. Harris, ‘Primer on binary logistic regression’, *Fam. Med. Community Health*, vol. 9, no. Suppl 1, p. e001290, Dec. 2021, doi: 10.1136/fmch-2021-001290.

[78] L. P. Blything and R. Blything, ‘The psychological (ab)use of timeouts in professional tennis’, *Int. J. Perform. Anal. Sport*, pp. 1–19, Dec. 2024, doi: 10.1080/24748668.2024.2436271.

[79] M. Fritz and P. D. Berger, ‘Will anybody buy? Logistic regression’, in *Improving the User Experience Through Practical Data Analytics*, Elsevier, 2015, pp. 271–304. doi: 10.1016/B978-0-12-800635-1.00011-2.

[80] W. H. Finch, J. E. Bolin, and K. Kelley, ‘Multilevel Modeling Using R’, 2014.

[81] E. Herold and C. Breuer, ‘Does the Game Matter? Analyzing Sponsorship Effectiveness and Message Personalization in Sport Live Broadcasts’, *J. Sport Manag.*, vol. 37, no. 4, pp. 290–301, Jul. 2023, doi: 10.1123/jsm.2022-0206.

[82] J. Dabass, ‘Demystifying Logistic Regression: Why Feature Scaling Matters’. Accessed: Apr. 16, 2025. [Online]. Available: https://blog.venturemagazine.net/demystifying-logistic-regression-why-feature-scaling-matters-d0729eca43ce#:~:text=Logistic%20Regression-,In%20practice%2C%20feature%20scaling%20can%20significantly%20improve%20the%20performance%20of,applied%20before%20training%20the%20model.

[83] D. Berrar, ‘Performance Measures for Binary Classification’, in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 546–560. doi: 10.1016/B978-0-12-809633-8.20351-8.

[84] C. Strobl, J. Malley, and G. Tutz, ‘An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and

random forests.', *Psychol. Methods*, vol. 14, no. 4, pp. 323–348, Dec. 2009, doi: 10.1037/a0016973.

[85] B. B. Hanberry, 'Practical guide for retaining correlated climate variables and unthinned samples in species distribution modeling, using random forests', *Ecol. Inform.*, vol. 79, p. 102406, Mar. 2024, doi: 10.1016/j.ecoinf.2023.102406.

[86] H. Alsagri and M. Ykhlef, 'Quantifying Feature Importance for Detecting Depression using Random Forest', *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 5, 2020, doi: 10.14569/IJACSA.2020.0110577.

[87] D. M. Vistro, F. Rasheed, and L. G. David, 'The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics', vol. 8, no. 09, 2019.

[88] A. Sarkar, 'Identification and Selection of Rising Performers in T20 Cricket Using Machine Learning Algorithms', 2023.

[89] C. Wang, 'Optimization of sports effect evaluation technology from random forest algorithm and elastic network algorithm', *PLOS ONE*, vol. 18, no. 10, p. e0292557, Oct. 2023, doi: 10.1371/journal.pone.0292557.

[90] Z. Gao and A. Kowalczyk, 'Random forest model identifies serve strength as a key predictor of tennis match outcome', *J. Sports Anal.*, vol. 7, no. 4, pp. 255–262, Dec. 2021, doi: 10.3233/JSA-200515.

[91] H. Wang, Q. Lu, and C. Chen, 'Prediction of Tennis Match Based on Random Forest and PSO-SVM', in *Proceedings of the 2024 International Conference on Smart City and Information System*, Kuala Lumpur Malaysia: ACM, May 2024, pp. 556–560. doi: 10.1145/3685088.3685185.

[92] Z. Shen, J. Lei, and Y. You, 'Tennis Player Winning Prediction Based on Grid-Search Random Forest Algorithm', in *2024 International Conference on Telecommunications and Power Electronics (TELEPE)*, Frankfurt, Germany: IEEE, May 2024, pp. 560–563. doi: 10.1109/TELEPE64216.2024.00106.

[93] C. Pham, K. Bufi, and J. Westerlund, 'Predicting Tennis Match Results Using Classification Methods', 2023.

[94] M. McHugh, 'Interrater reliability: the kappa statistic', *Biochem Med Zagreb*, vol. 22, no. 3, pp. 276–282, Oct. 2012.

[95] A. Agresti, 'Introduction to Categorical Data Analysis', *John Wiley Sons*, 2018.

[96] V. P. D. F. Melonio, M. S. Aoki, A. F. S. Arruda, D. Souza, C. D. Capitani, and A. Moreira, 'Analysis of serve and serve return on different surfaces in elite tennis

players', *Rev. Bras. Cineantropometria Desempenho Hum.*, vol. 23, p. e76603, 2021, doi: 10.1590/1980-0037.2021v23e76603.

[97] Z. Jeffery, 'How Does The Weather and Climate Affect Tennis?', Topspin Pro. Accessed: May 30, 2025. [Online]. Available: https://topspinpro.com/blog/how-does-the-weather-and-climate-affect-tennis/

[98] J. Ungvarsky, 'Yerkes-Dodson law', 2020.

[99] J. V. Tu, 'Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes', *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, Nov. 1996, doi: 10.1016/S0895-4356(96)00002-9.

[100] V. Candila and L. Palazzo, 'Neural Networks and Betting Strategies for Tennis', *Risks*, vol. 8, no. 3, p. 68, Jun. 2020, doi: 10.3390/risks8030068.

[101] J. Block and A. M. Kremen, 'Ego-Resiliency Scale'. 1996. doi: 10.1037/t01072-000.

[102] J. G. Eisenhauer, 'Explanatory Power and Statistical Significance', *Teach. Stat.*, vol. 31, no. 2, pp. 42–46, Jun. 2009, doi: 10.1111/j.1467-9639.2009.00364.x.

[103] P. Young, 'Impact of Label Noise on Model Generalization', Nov. 2024.

[104] J. P. Fuentes-García, M. Crespo, R. Martínez-Gallego, and S. Villafaina, 'Impact of match-induced pressure on HRV of junior tennis players', *Physiol. Behav.*, vol. 252, p. 113836, Aug. 2022, doi: 10.1016/j.physbeh.2022.113836.

[105] S. Pineda-Hernández, 'Playing under pressure: EEG monitoring of activation in professional tennis players', *Physiol. Behav.*, vol. 247, p. 113723, Apr. 2022, doi: 10.1016/j.physbeh.2022.113723.

[106] S. Pineda-Hernández, 'How to play under pressure: EEG monitoring of mental activation training in a professional tennis player', *Physiol. Behav.*, vol. 250, p. 113784, Jun. 2022, doi: 10.1016/j.physbeh.2022.113784.

[107] A. D. Crosswell and K. G. Lockwood, 'Best practices for stress measurement: How to measure psychological stress in health research', *Health Psychol. Open*, vol. 7, no. 2, p. 2055102920933072, Jul. 2020, doi: 10.1177/2055102920933072.

[108] Z. Bai and X. Bai, 'Sports Big Data: Management, Analysis, Applications, and Challenges', *Complexity*, vol. 2021, no. 1, p. 6676297, Jan. 2021, doi: 10.1155/2021/6676297.

[109] T. Endo, H. Sekiya, and C. Raima, 'Psychological pressure on athletes during matches and practices', *Asian J. Sport Exerc. Psychol.*, vol. 3, no. 3, pp. 161–170, Dec. 2023, doi: 10.1016/j.ajsep.2023.07.002.

[110] I. Prieto-Lage, A. Paramés-González, D. Torres-Santos, J. C. Argibay-González, X. Reguera-López-de-la-Osa, and A. Gutiérrez-Santiago, 'Match analysis and probability of winning a point in elite men's singles tennis', *PLOS ONE*, vol. 18, no. 9, p. e0286076, Sep. 2023, doi: 10.1371/journal.pone.0286076.

[111] R. O'Brien and H. Ishwaran, 'A random forests quantile classifier for class imbalanced data', *Pattern Recognit.*, vol. 90, pp. 232–249, Jun. 2019, doi: 10.1016/j.patcog.2019.01.036.

[112] A. Pandey, B. A. Shivaji, M. Acharya, and K. K. Mohbey, 'Mitigating class imbalance in heart disease detection with machine learning', *Multimed. Tools Appl.*, Jun. 2024, doi: 10.1007/s11042-024-19705-8.

[113] K. Suresh and S. Chandrashekara, 'Sample size estimation and power analysis for clinical research studies', *J. Hum. Reprod. Sci.*, vol. 5, no. 1, p. 7, 2012, doi: 10.4103/0974-1208.97779.

[114] K. Davids, 'Ecological Validity in Understanding Sport Performance: Some Problems of Definition', *Quest*, vol. 40, no. 2, pp. 126–136, Aug. 1988, doi: 10.1080/00336297.1988.10483894.

[115] S. Boué, M. Byrne, A. W. Hayes, J. Hoeng, and M. C. Peitsch, 'Embracing Transparency Through Data Sharing', *Int. J. Toxicol.*, vol. 37, no. 6, pp. 466–471, Nov. 2018, doi: 10.1177/1091581818803880.

[116] M. Y. Balban *et al.*, 'Brief structured respiration practices enhance mood and reduce physiological arousal', *Cell Rep. Med.*, vol. 4, no. 1, 2023.

[117] Y. A. Balk, M. A. Adriaanse, D. T. D. De Ridder, and C. Evers, 'Coping Under Pressure: Employing Emotion Regulation Strategies to Enhance Performance Under Pressure', *J. Sport Exerc. Psychol.*, vol. 35, no. 4, pp. 408–418, Aug. 2013, doi: 10.1123/jsep.35.4.408.

[118] L. J. Moore, S. J. Vine, M. R. Wilson, and P. Freeman, 'Reappraising Threat: How to Optimize Performance Under Pressure', *J. Sport Exerc. Psychol.*, vol. 37, no. 3, pp. 339–343, Jun. 2015, doi: 10.1123/jsep.2014-0186.

[119] Q. Wang *et al.*, 'Analysing tie-break performance of professional tennis players at Grand Slam matches', *J. Sports Sci.*, pp. 1–9, Aug. 2024, doi: 10.1080/02640414.2024.2384244.

[120] M. Veenman, A. M. Stefan, and J. M. Haaf, 'Bayesian hierarchical modeling: an introduction and reassessment', *Behav. Res. Methods*, vol. 56, no. 5, pp. 4600–4631, Sep. 2023, doi: 10.3758/s13428-023-02204-3.

[121] C. Van Landingham, W. G. Fuller, and R. A. Schoof, ‘The effect of confounding variables in studies of lead exposure and IQ’, *Crit. Rev. Toxicol.*, vol. 50, no. 9, pp. 815–825, Oct. 2020, doi: 10.1080/10408444.2020.1842851.

[122] ITF Council, ‘Rules of tennis 2025 amendments’, Jan. 2025. Accessed: Jun. 05, 2025. [Online]. Available: https://www.itftennis.com/media/7224/2025-amendments-to-the-rules-of-tennis-and-beach-tennis.pdf

[123] BBC, ‘Off-court coaching to be allowed in tennis from 2025’. Accessed: Jun. 05, 2025. [Online]. Available: https://www.bbc.co.uk/sport/tennis/articles/cglklyl048no

Appendices

Appendix A – Descriptive Statistics

TABLE AI.

BREAKDOWN OF MATCHES BY COURT SURFACE AND ROUND

Surface	Total	Final	Semi Finals	Quarter Finals	Round of 16	Round of 32	Round of 64	Round of 128
Clay	477	13	30	52	95	178	75	34
Grass	158	3	7	16	33	56	14	29
Hard	617	17	25	61	107	194	132	81

TABLE AII.

BREAK POINT WIN RATES BY COURT SURFACE

Surface	Total Break Points	Break Points Won	Win Rate
Clay	6,574	2,745	0.418
Hard	8,193	3,381	0.413
Grass	2,036	820	0.403

TABLE AIII.

BREAK POINT WIN RATES BY ROUND

Round	Total Break Points	Break Points Won	Win Rate
Round of 128	2,433	1,025	0.421
Round of 64	3,418	1,384	0.405
Round of 32	5,311	2,239	0.422
Round of 16	2,844	1,178	0.414
Quarter Finals	1,613	659	0.409
Semi Finals	778	303	0.389
Final	406	158	0.389

TABLE AIV.

BREAK POINT WIN RATES BY RANKING CATEGORY

Ranking Category	Total Break Points	Break Points Won	Win Rate
51–100	5,385	2,300	0.427
31–50	3,108	1,301	0.419
101-200	2,338	959	0.410
Top 10	1,755	719	0.410
200+	1,067	428	0.401
11–30	3,150	1,239	0.393

Appendix B – Confusion Matrices

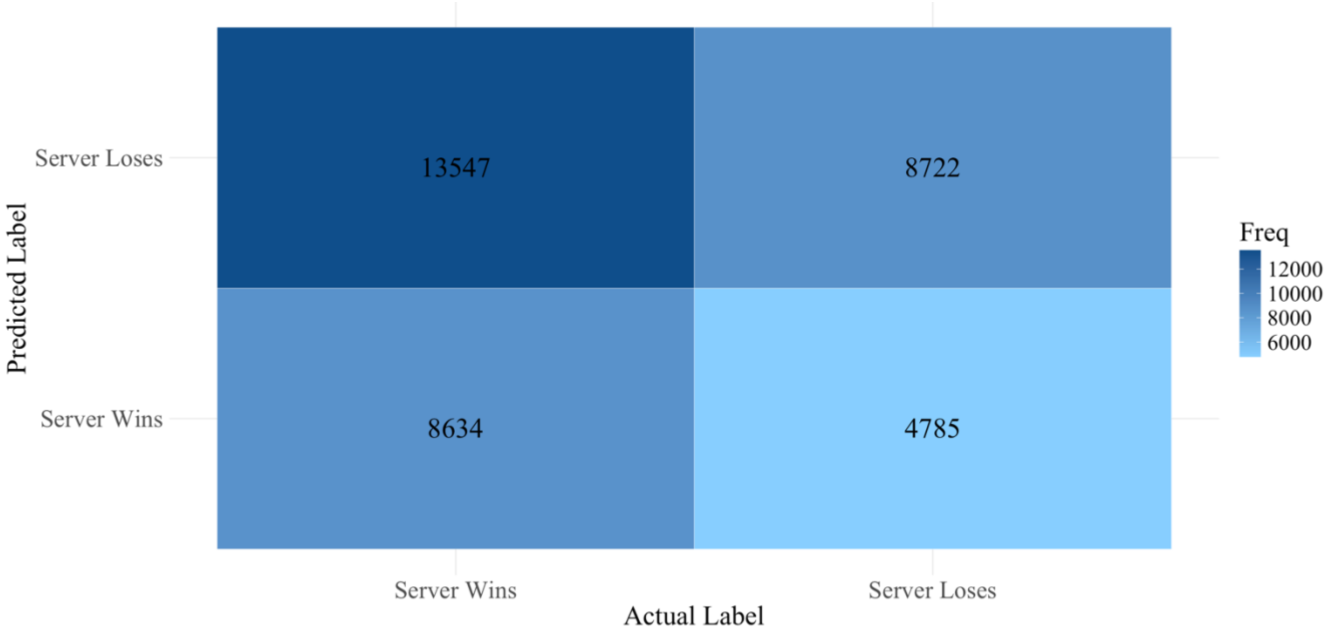

Fig. B1. Confusion Matrix for Model 1 predictions on test data

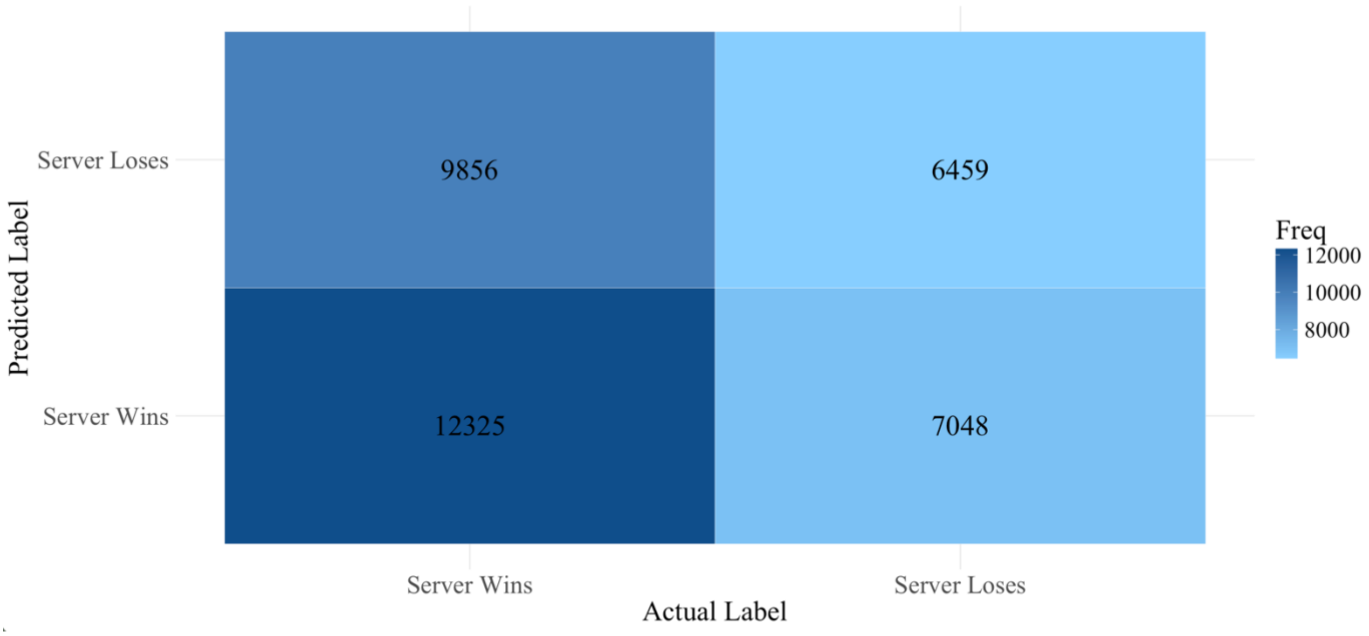

Fig. B2. Confusion Matrix for Model 2 predictions on test data

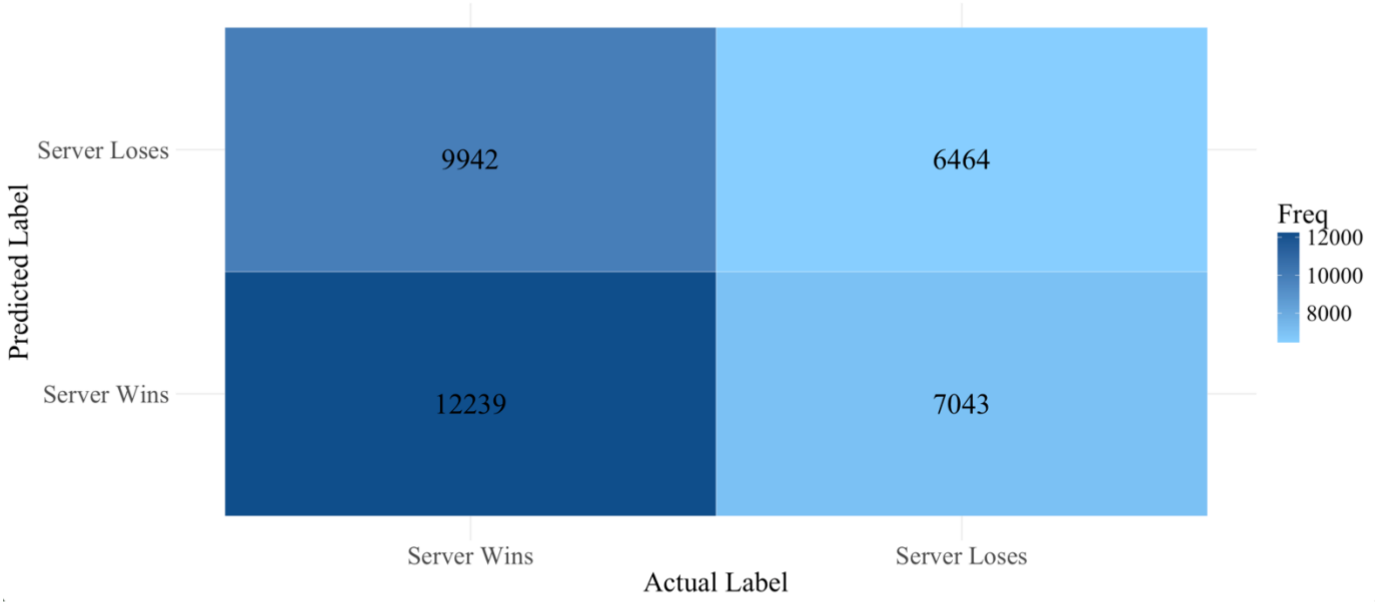

Fig. B3. Confusion Matrix for Model 3 predictions on test data

Fig. B4. Confusion Matrix for Model 4 predictions on test data

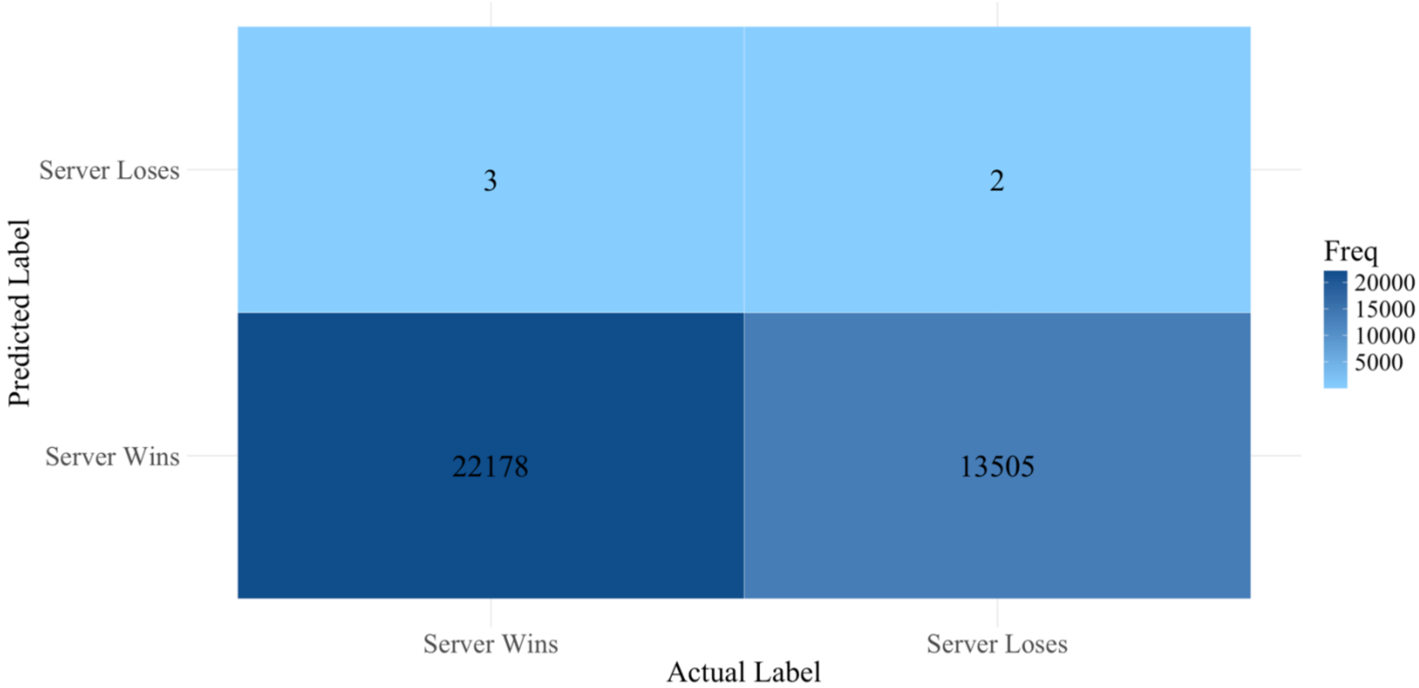

Fig. B5. Confusion Matrix for Model 5 predictions on test data

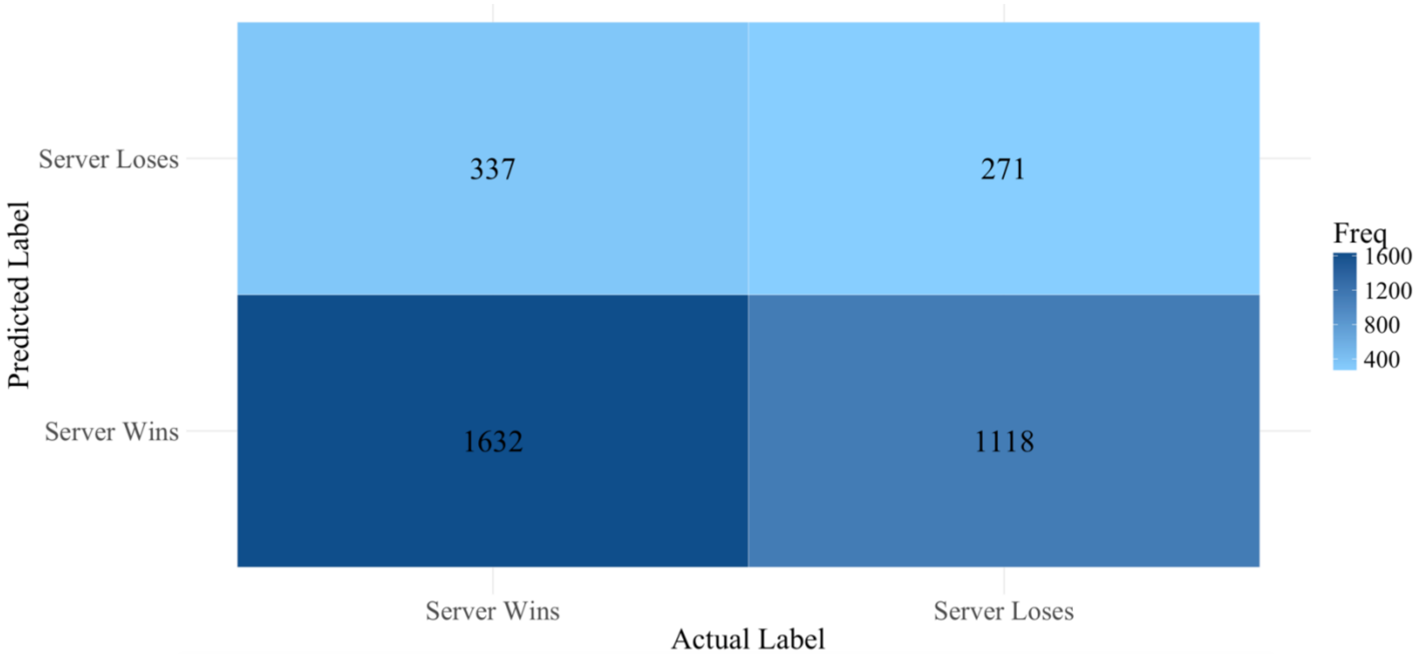

Fig. B6. Confusion Matrix for GLMM follow-up analysis on isolated break points

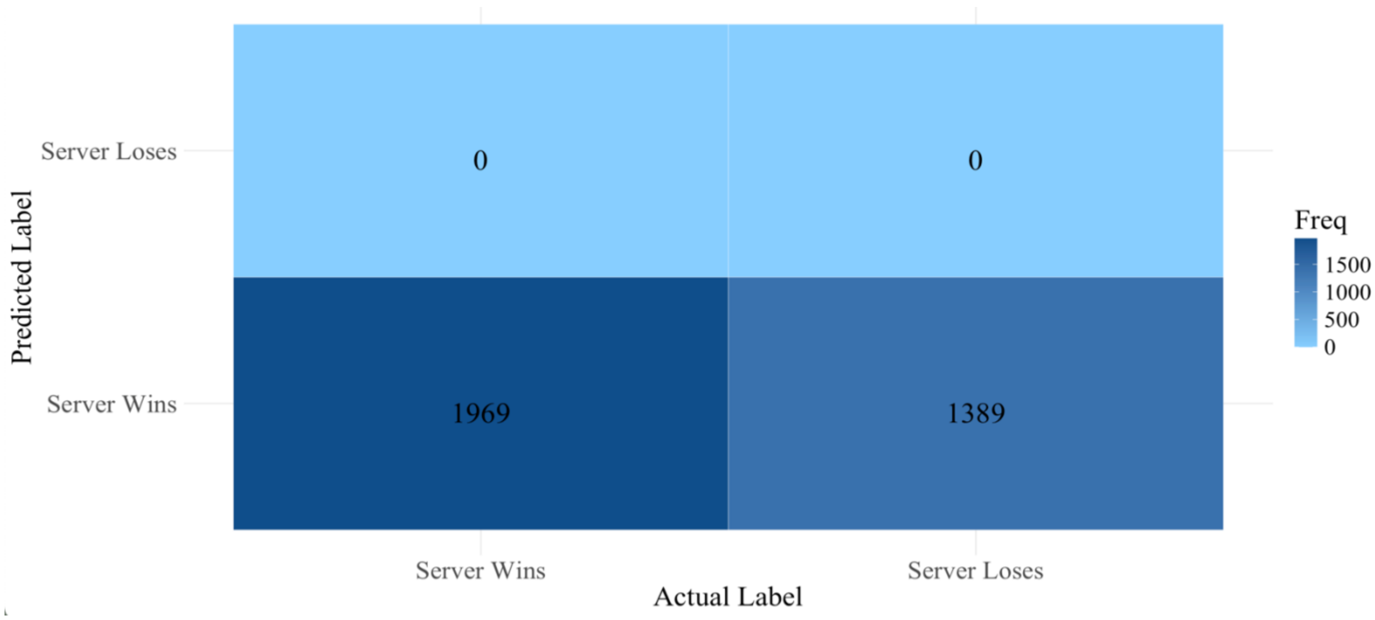

Fig. B7. Confusion Matrix for RF follow-up analysis on isolated break points

Appendix C – Exploratory Analysis

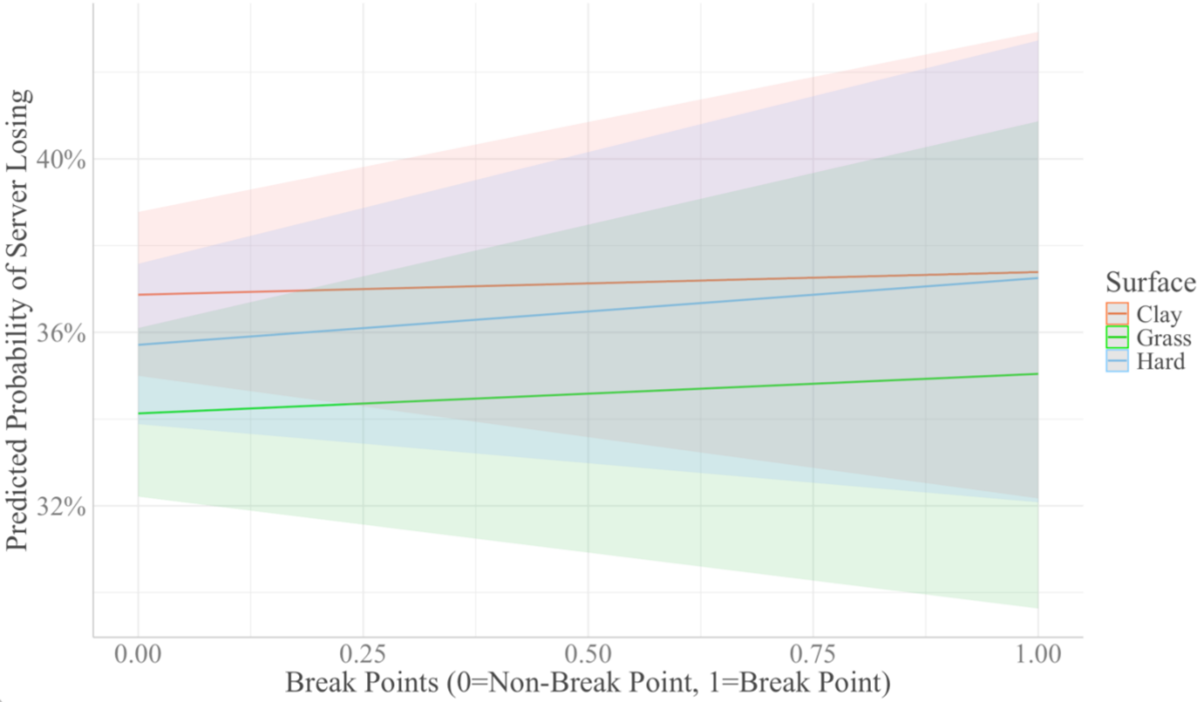

Fig. C1. Marginal effects of break points by surface