Momentum in Tennis

Logan Schindler and Tyler Teichmiller
MCS358

August 13, 2024

1 Abstract

Momentum in tennis is both difficult to measure and not well-defined. This paper first defines momentum within the context of tennis, then explores ways to use neural networks to predict the win probability of players and the flow of play in a tennis match. Using the 2023 Wimbledon Gentlemen's point-by-point data, we trained these neural networks and created a definition of momentum in the context of tennis to measure the phenomenon. The first neural network predicts the win probability difference for two players with a R^2 value of .9342, and a second neural network accurately predicts the winner of a point 99.135% of the time. Using our definition of momentum, we analyze its accuracy on the 2023 Wimbledon Gentelmen's Final match, and conclude that momentum is not random, can be quantified, and can positively indicate player performance during a match.

2 Introduction

Novak Djokovic entered the 2023 Wimbledon Gentlemen's final as the four-time reigning champion, looking to match the record for Wimbledon victories set by Roger Federer and be the first man since 1969 to complete the Grand Slam in a calendar year [Pre23]. Standing in his way was rising star Carlos Alcaraz of Spain. The entire match seemed to be a back-and-forth battle. It saw Djokovic take a commanding 5-0 lead in set one before Alcaraz powered back to win sets two and three by a score of 7-6 and 6-1, respectively. Djokovic, on the brink of elimination, forced a tiebreaker set by winning set four 6-3. Finally, Alacaraz came out on top winning the final set 6-4 and denying Djokovic a fifth-straight Wimbledon win, a calendar year Grand Slam, and one of the greatest seasons in tennis history [23a].

Wimbledon is the oldest and most prestigious of the Grand Slam tournaments, a single-elimination 128-person tournament where the best players in the world play on grass courts. The grass courts used for matches make Wimbledon unique from other Grand Slam majors. Grass courts have less friction, so the ball maintains its high speeds, making the style of play more advantageous for players with powerful serves [Nag22]. To win the tournament, a player must win all seven matches to be crowned the champion.

At Wimbledon, a player wins a match by winning three of five sets. To win a set, you have to win 6 games, but you have to win by at least 2 games until the set is tied $6-6$. Then, a tiebreaker is played. In the tiebreaker, players alternate serving two points until someone scores at least 7 points and wins by 2 points. And to win a game, a best of seven points, win by 2, is played to determine the winner.

Since serving is one of the biggest factors in deciding who wins a point, rules are established to delegate who should serve. At the beginning of the match, it is randomly chosen that one person will decide whether they want to start serving the first game. The two players will alternate serving games for the remainder of the set. Once the set has ended, the player who didn't serve in the first game will now serve the first game of the new set. This way, the advantage of increased serving speeds caused by the grass courts is evenly distributed between the two players.

One sports phenomenon that athletes, coaches, and spectators claim to observe during back-and-forth competitions such as the 2023 Wimbledon Gentlemen's Final is momentum [23b]. Momentum, while being a precisely defined and measurable phenomenon in physics, is not defined with the same ease in the context of sports. Some skeptics claim that momentum is not a real, observable, or quantifiable occurrence in sport. Instead, they claim that it is "one of the most meaningless bits of jargon in coachspeak"[Hal21]. Skeptics claim that momentum is a scapegoat concept that coaches use to deflect blame from their coaching staff or their players to the 'sporting gods.'

Proponents claim that not only is momentum in sports a real phenomenon, but that it can be both measured and described. By training a neural network on 10 seasons worth of play-by-play data from the National Football League, University of Wisconsin-Milwaukee Professor and Researcher Paul Roebber was able to both define momentum and show that swings in play do not always come down to random chance [Ott22].

This paper looks to define momentum in the context of sports; specifically, in tennis at the Wimbledon Championships. We use a definition similar to that of Roebber's definition: a change in the difference in win probability over two consecutive possessions. In tennis, we define a possession to be one game. Thus, each player will have the opportunity to be both the server and the returner in each calculation of momentum.

This paper also looks to model the flow of play in a tennis match as points occur. This way, it can identify which player performs better during a match. We model the flow of play in two similar but distinct ways. First, we will use a Decision Tree neural network to create a model that predicts the winner of a point based on different actions that occur during the point. This will help identify who might perform better in the immediate future. Second, we will use a Regression neural network to predict each player's win probability difference as the match progresses. This allows us to measure the flow of the play and makes it easy to compare player performances with each other.

In Section 3, we explore the dataset used to assist our mathematical analysis of momentum and the flow of play. In Section 4, we explore the assumptions for our models before describing them in Section 5. In Section 6, we analyze the models and apply it to certain games from the 2023 Wimbledon Championships tournament. Finally, in Section 7, we discuss limitations with our models and give recommendations for future work.

3 Data

During the course of the 2023 Wimbledon tournament, different metrics were recorded during every point played after the first 2 rounds of the tournament. This data is in Appendix A in the subsection Wimbledon Data, and contains information on 7284 individual points over 31 completed matches. This corresponds to every Gentlemen's singles match after the second round of 2023 Wimbledon. This produces a dataset where the match's score can be analyzed during any portion of the match. For example, during the round 3 match between Carlos Alcaraz and Nicolas Jarry, we can analyze the individual point when the score was (2-1, 0-3, 0,0). This shows that Carlos Alcaraz has won 2 sets to Nicolas Jarry's 1, Alcaraz is losing 0 games to Jarry's 3, and a new game has started since the score is 0 to 0.

Tracking each point allows for predictions and adjustments to be made

Summary Statistics for Win Probability Difference							
Minimum	Q1	Median	Q3	Maximum	Mean	St Dev	n
-0.9999	-0.4372	0.0357	0.4397	1	0.0166	0.5564	7284

Table 1: Summary Statistics for Win Probability Difference for all 7284 points in wimbledon_serve_win_prob.csv

during the match. In addition to the match's score, data such as the distance run during the point, the speed of the serve, and where serves were hit, among other variables, were collected and recorded. The original dataset contains information on 46 different variables. Additional data from the ATP tour website was also collected and recorded to use in our models [ATP]. The ATP Tour is the global governing body for men's tennis and is the governing body that runs Wimbledon. For each player included in the dataset, we collected their 2022 match statistics from when they were playing on grass courts. These statistics were their first serve win percentage, first serve in bounds percentage, second serve win percentage, service point win percentage, and return point win percentage. This way, their performance can be measured on grass courts leading up to Wimbledon. There were some players who did not have any metrics on grass courts, so we had to take measurements from other time periods, such as their data from 2021 or 2023 instead. After these additions, the final dataset used in our models has information on 60 total variables.

The data attributes in Table 2 are used to create and train the models in our paper. For a complete list of variables in the dataset, please see Appendix A. The first model, Win Probability Model, predicts the difference in win probability between player 1 and player 2. Summary statistics for win probability difference can be found in Table 1. The second model, Point Winner Model, predicts the value of the variable "Point Victor." Point Victor is a binary response variable with values 1 and 2. The value 1 means that player 1 won the point, and the value 2 means that player 2 won the point. Of all 31 matches in the dataset, 3718/7284, or 51.043% of all points were won by player 1, and the remaining 3566/7284, or 48.957% of points were won by player 2.

4 Model Assumptions

4.1 Win Probability Model and Point Winner Model

The model assumptions for the Win Probability Model and the Point Winner Model are the same. First, we assume that all points played are at the 2023 Wimbledon Championships beyond the second round of the tournament. Last, we assume that each player has a constant first serve win percentage, first serve in bounds percentage, second serve win percentage, service point win percentage, and return point win percentage throughout the match.

5 Model Description

5.1 Win Probability Program

The Win Probability Program has been adapted from the Python files on Jeff Sackmann's Github page, which uses absorbing Markov chains to calculate win probabilities. The original Python program was written in Python2, which is outdated and not supported. So, changes had to be made to make the Python files compile and run on Python3 interpreters. The first change was changing some of the division symbols from '/' to '//'. In Python2, the interpreter would automatically perform integer division using the '/' symbol, but it no longer does that. So, some of the single slashes were changed to double slashes to perform integer division when specified.

The next change was adjusting the outputs of certain functions. Some functions output a tuple with an integer as the first item and a dictionary as the second item, but other functions that relied on that output only needed the integer output. So, changes were made to which types of outputs certain functions would return. These changes allowed the program to function properly with our data and Python3 interpreters.

The program has a match win probability function that takes in multiple inputs. Those inputs are a player's service win percentage, the same player's return win percentage, the score within a game for both players, the number of games won for each player, the number of sets won for each player, and whether the match was a best of 3 or best of 5 sets match. It will then calculate the probability of a player winning the match based on the current match score.

The match win probability function works since the Win Probability Program has many functions to calculate the win percentage for smaller parts of the match. The match win probability function first calls a function to calculate the probability of winning and losing the current set. Then, it will calculate the probability of winning the match based on those two scenarios and sum them together to get the total win probability of winning the match.

However, the current set win probability function also uses other functions to calculate its result. First, it calls a function to find the probability of winning and losing the current game. Then, it calculates the probability of winning the set based on those two scenarios and sums them together to get the total win probability of winning the set, which is then used in the match win probability function.

The main drawback of this program is that it calculates the win percentage using only a single player's service and return metrics. This creates an issue when two elite players are competing against each other since they will both have good metrics. So, if the win probability program was used for both players at the same point, then they might both have win probabilities above 50%. To solve this problem, we will look at a new metric: the win probability difference between the two players. This way, the player with the better metrics will have the advantage over the other player.

The program can now calculate the win probability difference for each point played in the Wimbledon dataset. This is done by writing a script that calls the match win probability function for both players for every point using each players respective serving statistics discussed previously. Those two win probabilities are then used to calculate the win probability difference and are added to the dataset as a new variable. Other models will use the win probability difference metric to track the flow of the match and changes in momentum during the match.

For example, the first match in our dataset is between Carlos Alcaraz and Nicolas Jarry. Since Alcaraz began the match by serving, his win probability that corresponds to a service point win percentage of 69% and a return point win percentage of 36% is 77.6%. However, Jarry's service point win percentage is 70%, and his return point win percentage is also 36%. This leads to a win probability of 81.7% instead of the expected 22.4% if the probabilities of winning between the two players summed to one. Taking the difference in win probability allows us to see which player is performing better at what time on a scale that cannot be greater than one in absolute value at any time. Note that this value is calculated by taking $p_1 - p_2$, where

p_i is the win probability of player i. Using the above example, the difference between Alcaraz's and Jarry's win probabilities is -0.041, meaning Jarry has a higher chance of winning by 4.1%. The value is negative because Jarry is Player 2.

5.2 Win Probability Model

The Win Probability Model utilizes the win probabilities from the Win Probability Program. For each point in the original dataset, the win probability was calculated using the probability that the server wins a service point, the probability that the server wins a return point, the current score within the game, the current score within the set, and the current score within the match.

A neural network Regressor was then supervised-trained on the first 25 matches of the 31 match dataset for the purpose of regression. This split was chosen because the first 25 matches account for roughly 80% of the total number of points played in the dataset. The left column of Figure 2 shows the features that were used to train the Regressor. These features were chosen because they have no direct correlation to the response variable win probability difference. It was trained to predict the win probability difference described above. A Regressor was chosen because the values being predicted are continuous, not categorical. Finally, it was trained on the data in chronological order using a time-series split. This ensures that points are not seen as standalone points, but as points that are in the broader context of the game, set, and match that contains them.

5.3 Point Winner Model

The Point Winner Model was trained on the same dataset as the Win Probability Model. It was trained on a smaller number of features that can be found in the right column of Figure 2. Some features from the Win Probability Model were not included as features in the Point Winner Model due to their direct correlation to points scored. For example, if we know player 1 had two points in the game at the last point, and now has three points, we know player 1 must have won the point. Similar arguments can be made about games won and sets won, and can be generalized to player 2's values for these features.

Features	
Win Probability Model	Point Winner Model
Server	Server
Serve number	Serve Number
Distance Ran by Player 1	Distance Ran by Player 1
Distance Ran by Player 2	Distance Ran by Player 2
Rally Count	Rally Count
Proportion of Player 1's First Serves that are In Bounds	Proportion of Player 1's First Serves that are In Bounds
Proportion of Player 1's First Serves That Player 1 Wins	Proportion of Player 1's First Serves That Player 1 Wins
Proportion of Player 1's Second Serves That Player 1 Wins	Proportion of Player 1's Second Serves That Player 1 Wins
Proportion of Player 1's Service Points that Player 1 Wins	Proportion of Player 1's Service Points that Player 1 Wins
Proportion of Player 1's Return Points that Player 1 Wins	Proportion of Player 1's Return Points that Player 1 Wins
Proportion of Player 2's First Serves that are In Bounds	Proportion of Player 2's First Serves that are In Bounds
Proportion of Player 2's First Serves That Player 2 Wins	Proportion of Player 2's First Serves That Player 2 Wins
Proportion of Player 2's Second Serves That Player 2 Wins	Proportion of Player 2's Second Serves That Player 2 Wins
Proportion of Player 2's Service Points that Player 2 Wins	Proportion of Player 2's Service Points that Player 2 Wins
Proportion of Player 2's Return Points that Player 2 Wins	Proportion of Player 2's Return Points that Player 2 Wins
Number of Sets Player 1 Has Won	
Number of Games Player 1 Has Won	
Player 1's Score In Current Game	
Number of Sets Player 2 Has Won	
Number of Games Player 2 Has Won	
Player 2's Score In Current Game	

Table 2: All features used to train the Win Probability Model and Point Winner Model

The Point Winner Model was created by supervised-training a neural network Decision Tree for classification instead of a neural network Regressor. This is because the target, the variable "Point Winner", is a binary categorical variable with options 1 and 2, corresponding to Player 1 winning the point and Player 2 winning the point, respectively. A Decision Tree neural network is more appropriate to use to predict categorical targets than a Regressor is. The same training and testing split as the Win Probability Model was used, where the first 25 matches were used in the training set and the final 6 matches were used in the testing set.

5.4 Momentum

We will call the final section the Momentum section. For the last match in the dataset, which was the Wimbledon Final between Alcaraz and Djokovic, a statistic called momentum was calculated. We define it as a change in the win probability difference over two consecutive possessions. In tennis, we define a possession as one game. Thus, each player will have the opportunity to be both the server and the returner in each calculation of momentum. We used player 1's value of change in win probability to calculate this statistic. Similar to before, positive values of momentum will show momentum in favor of player 1 (negative momentum for player 2), and negative values indicate momentum in favor of player 2 (negative momentum for player 1).

6 Model Analysis

6.1 Win Probability Model

The Win Probability Model allows for the flow of play in a match to be measured and helps to quantify who is performing at a higher level. After training the Win Probability model, we measured its ability to accurately predict the difference in win probability on our testing set. This was evaluated by calculating and interpreting the R^2 value for the model, as well as the average squared distance from the predicted value to the actual value of the difference in win probability. Mathematically, the average squared distance can be explained by the expression

$$\sum_{i=0}^{n} \frac{(Actual_i - Predicted_i)^2}{n}$$

where $n = 1388$, the number of points played in the last 6 games at Wimbledon 2023, and $Actual_i$ and $Predicted_i$ are the Actual and Predicted value of the difference in win probability for the point number i, respectively.

The model has an R^2 value of 0.9342, which means that 93.42% of the variability in the difference in win probability is explained by the model. The average squared distance calculated for the final 6 games in the dataset is 0.01807.

We are interested in how this model predicts the 2023 Wimbledon Final match between Carlos Alcaraz and Novak Djokovic. Figure 1 shows the actual and predicted values for the difference in win probability using our Win Probability Model. Positive values indicate a higher win probability for Alcaraz, while negative values indicate a higher win probability for Djokovic. This is because Alcaraz is player 1, and Djokovic is player 2. The Win Probability Model initially overestimated the win probability difference in favor of Djokovic before returning to follow the actual line between points 75 and 150. Between points 150 and 250, the model slightly overestimates the win probability in favor of Alcaraz. Finally, the line accurately predicts the win probability difference for the final 90 points from point 250 to point 340. In this specific match, the average squared distance is 0.01659. While the win probability model is not perfect, it can model the flow of play in the Alcaraz versus Djokovic match and is able to give insights into who is performing at a higher level in the match by looking at the win probability difference.

One limitation to the Win Probability Model can be observed in Figure 2, which models the 2023 Wimbledon quarterfinal matchup between Andrey Rublev and Djokovic. The domain of values for the difference in win probability is only the interval $[-1, 1]$. This is because each player can at most have a 100% chance of winning the match and at least a 0% chance of winning the match, meaning their difference would be either -1 or 1. However, our model predicts values outside of this domain. In Figure 2, Djokovic has a much higher probability of winning the match than his opponent Andrey Rublev. This leads to the actual difference in win probability values being closer to -1 because Djokovic is Player 2. Our model often overestimates this probability in favor of Djokovic. At times, the model has a predicted value as low as -1.506, corresponding to Djokovic having a 150.6% win probability advantage over Rublev. Since this prediction is outside of the domain, this is a limitation of the model.

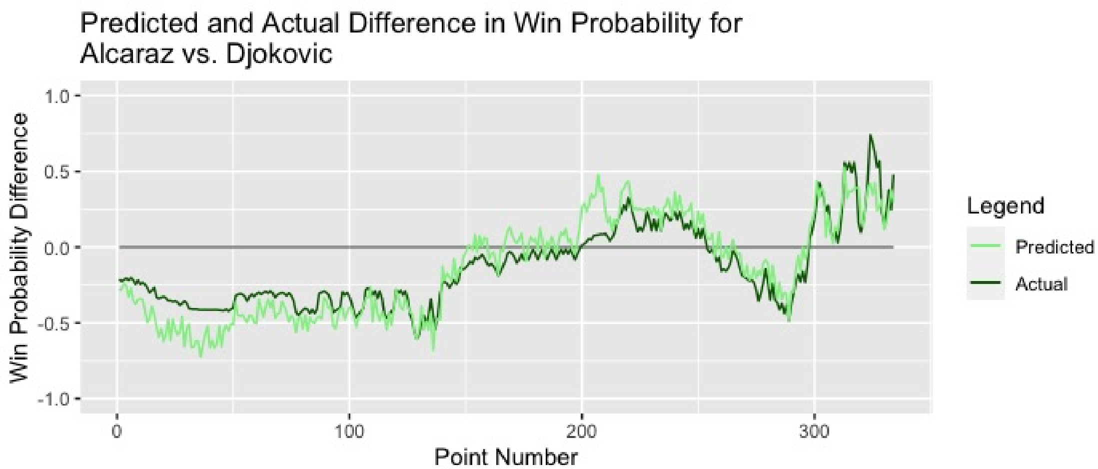

Figure 1: Predicted and Actual Win Probability Difference for the 2023 Wimbledon Final, Alcaraz vs. Djokovic.

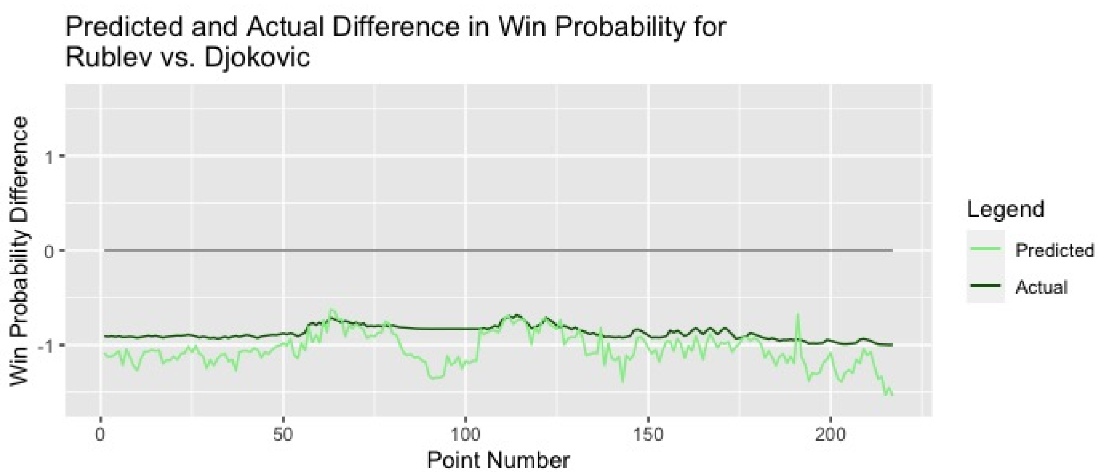

Figure 2: Predicted and Actual Win Probability Difference for the 2023 Wimbledon Quarterfinal, Rublev vs. Djokovic.

6.2 Point Winner Model

Using the Point Winner Model, we can make predictions about player performance on a point-by-point basis. After training the Point Winner Model, we similarly tested its accuracy, predicting the point winner for each individual point in the last six games of Wimbledon 2023. Overall, the model has an accuracy of 99.135%. This means that 99.135% of point victors in the last six games were accurately predicted. Only 12 out of the 1388 total points played in these games were predicted inaccurately. This model has a precision of 99.432%, meaning of the points that the model predicted player 1 would win, 99.432% actually were won by player 1. This model has a recall of 98.871%, meaning of points won by player 1, the model predicted that about 98.871% were won by player 1. A confusion matrix for this model's testing set can be found in figure 3. This confusion matrix shows a glimpse of which kinds of points were incorrectly predicted. It shows that of those points predicted to have been won by player 2, only 8 out of 683 points were won by player 1. Overall, this confusion matrix shows how accurate and precise this model is while predicting the final 6 matches of Wimbledon 2023.

While the entirety of the Decision Tree can be accessed and described, over 200 nodes are used to classify observations. Since Decision Trees prioritize reducing uncertainty while creating nodes, we know that the first node is the best predictor of which player wins the point. The first node in our Decision Tree splits based on the variable "server." If the server is greater than 1.5, then it splits to the left. If the server is less than or equal to 1.5, then it splits to the right. Since the variable "server" can only take the values of 1 and 2, this split is asking which player is serving. This aligns with our research into the subject matter. Coach Jon Carlson of the Gustavus Women's Tennis Team informed us that the server is usually the winner of the point in tennis, especially on a fast surface such as grass and at a high level of tennis like the Wimbledon Championships.

We are again interested in the performance of the Point Winner Model in the 2023 Wimbledon Final Match between Carlos Alcaraz and Novak Djokovic. A graph of the predicted flow of play and the actual flow of play in the entirety of this match can be found in Figure 4. The majority of the points in the match were correctly predicted. There are sections of play, like between points 63 and 73, or between points 116 and 135, where the two lines do not match. This is due to just 4 incorrect point predictions at point 63, 73, 116, and 135.

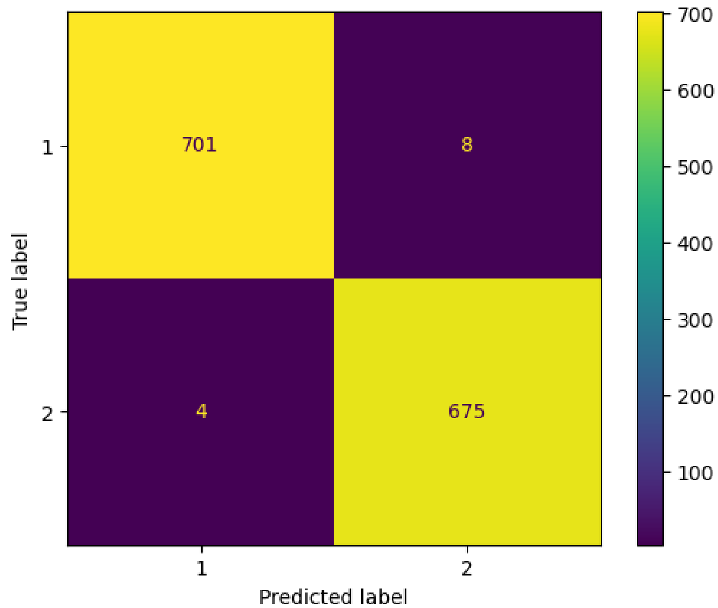

Figure 3: Confusion Matrix for Decision Tree Point Winner Model tested on the final six games of Wimbledon 2023.

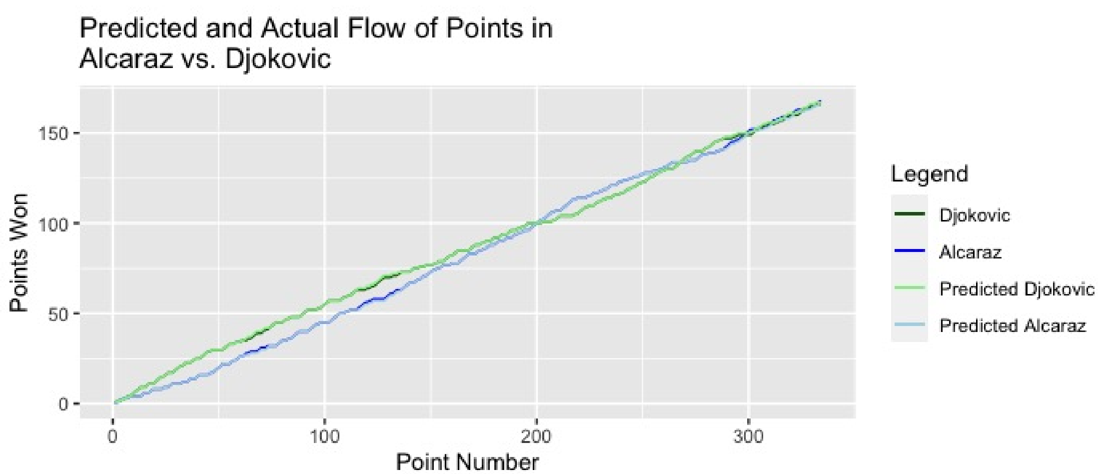

Figure 4: Predicted and Actual Flow of play for the 2023 Wimbledon Final between Alcaraz and Djokovic.

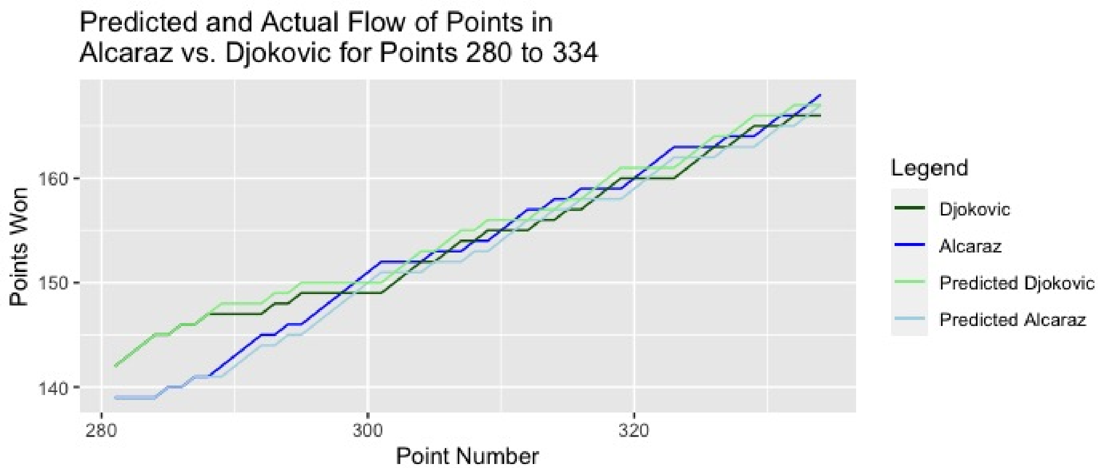

Figure 5: Predicted and Actual Flow of play for the 2023 Wimbledon Final between Alcaraz and Djokovic for final 54 points.

Then, the model is 100% accurate until point 289. A zoomed in look at the flow of play after point 289 can be found in Figure 5. A limitation to our Point Winner Model is that although it is very accurate in predicting the winner of the point, any incorrect prediction in a close match such as the 2023 Wimbledon Final can end up affecting the predicted winner of the overall match. In the predicted Alcaraz v. Djokovic match, each player finishes the match with 167 points. Another limitation is that the model utilizes variables to predict the winner of the point that, typically, we would not know if we didn't know the winner of the point. For example, the variable "p1_distance_ran" is a numerical variable that we would not know the final value of until the point was completed. Variables such as these may have increased the accuracy of our model.

6.3 Momentum

We again are interested in the measure of momentum in the match between Alcaraz and Djokovic to measure which player is performing better at a specific point in the match. Figure 6 shows our defined measure of momentum plotted for this match. The black horizontal lines are at 0.019 and -0.019. Momentum calculations between these two lines are considered to be random chance, according to Roebber [Ott22].

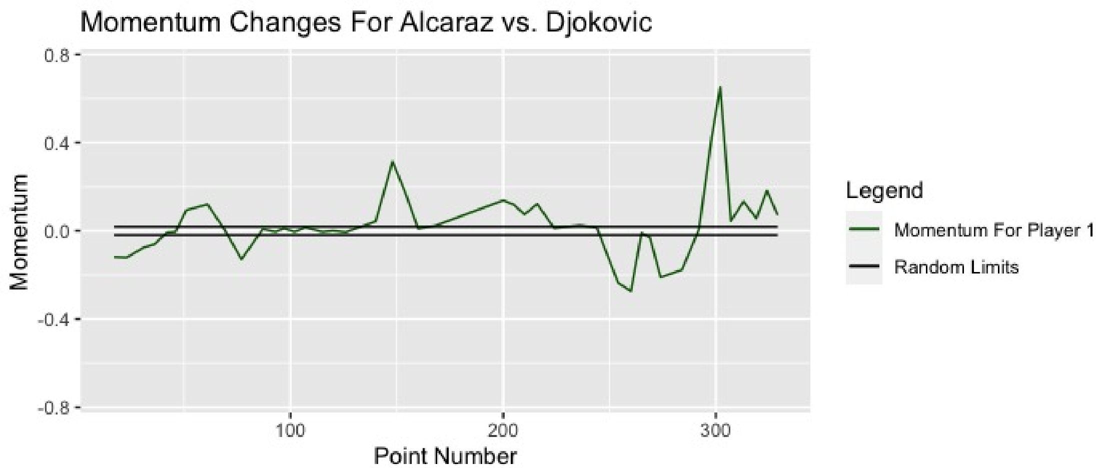

Figure 6: Momentum Changes for the 2023 Wimbledon Final between Alcaraz and Djokovic.

Momentum values outside of these values show momentum in favor of a specific player. Since Alcaraz is player 1 in this match, positive values of momentum correlate to Alcaraz performing better and controlling momentum, while negative values of momentum correlate to momentum in favor of Djokovic. Analysis of this graph shows a story similar to that described in the introduction. Djokovic wins his first set 6-1 at point number 46, which is when the momentum line begins to be random chance. After this, Alcaraz gains momentum by winning the second set in the tiebreaker between points 47 and 140, where at point 140, Alcaraz sees a sharp increase in momentum. Alcaraz holds the momentum through the third set, as he wins 6-1 between points 141 and 210. After this, Djokovic regains momentum during the fourth set, winning 6-3. With the game score of 1-1 in the fourth set, momentum begins to favor Alcaraz as he goes on to win the set 6-4 and the match 3 sets to 2. So, momentum as it is define in this paper, correlates to the events that occur during a match.

Overall, our momentum statistic identifies which player is performing better and quantifies how much better they are doing at any given point. Since momentum is defined as the change in win probability over the change of two possessions, variables that impact win percentage will also impact momentum.

7 Discussion and Conclusion

This paper has used data on 7284 individual points scored in the 2023 Wimbledon Championships in every Gentlemen's singles match after the second round to train two models. The first model, the Win Probability Model, utilized the Win Probability Program to calculate the probability that each player wins the match based on the current score of the match (game, set, and match scores) and their win percentages while serving and returning points. Then, the difference between the players' win probabilities was calculated. A Regressor neural network was then trained and fit on the first 25 matches in the dataset using a time series split that allows points to be not seen as standalone points, but as points within the broader context of the game, set, and match that contains them. This model has an R^2 value of 0.9342 and an average squared distance of 0.01807 for the final 6 games in the dataset, which were not used to train the model. With this model, we could measure the flow of play during a match, and by looking at the win probability difference, we could determine who is performing at a higher level during the match.

The second model, the Point Winner Model, used a Decision Tree Classifier neural network to predict the winner of an individual point. This model was trained using fewer features than the Win Probability Model due to the correlation between select features and the response variable Point Victor. Again, the first 25 matches of the dataset were used as training data using a time series split that allows points to be looked at in the broader scope of the game, set, and match that contains them. When testing this model on the final 6 unseen matches in the dataset, we see that the model correctly predicts the winner of the point 99.135% of the time. It also features a recall score of 98.871% and a precision of 99.432%. Overall, this model is incredibly good at predicting the winner of a specific point in the 2023 Wimbledon final 6 matches. Allowing the model to predict performance in the immediate future. The most important feature used to predict the point's winner was which player was serving. The vast majority of the point victors will be the serving player!

A variable called momentum was calculated for the final match of the dataset, which was the Wimbledon Final between Carlos Alcaraz and Novak Djokovic. When defining momentum, we used a similar definition to the one used by Paul Roebber in his research paper measuring momentum during football matches [RBd22]. We define it to be a change in the difference in win probability over two consecutive possessions, with possessions being one game. When graphing momentum as points are played in the match and then corroborating this information with accounts of the match between Alcaraz and Djokovic, we see that momentum under our definition does exist and does not come down to random chance.

Both models have specific limitations that were discussed in the Model Analysis sections of this paper. However, there are general limitations that have not yet been addressed, and that could lead to future work on the topic. First, neither model will apply to any tournament that isn't played on grass because our percentages taken were for grass matches only. Often, players are more effective on a specific court type. For example, a special court was made for a match between Roger Federer and Rafael Nadal in 2007 labeled "The Battle of the Surfaces" where half of the court was clay and half was grass. This was due to Federer's domination on grass courts and Nadal's domination on clay courts[Sok24].

Second, the Win Probability Program calculates win probability only based on one player's information. Future work could look at creating a better program that utilizes both players serving statistics, along with the current score of the match, to calculate win probability.

Third, not all players had data on their serving percentages on grass courts for the year 2022. Due to this, four players' statistics for years other than 2022 were used. Future work could utilize the serving percentages for the match up to the point played for both players to give a more accurate idea of how good each player is at serving and returning points.

Finally, neither model would apply to tournaments other than the Wimbledon Gentlemen's Singles tournament. This is because often, women tennis players have lower win percentages on service points than men [Rao20]. Additionally, doubles matches would not apply because our model only considers one player on each side of the net serving and returning, not two. Additional statistics would need to be collected, and adjustments to the Win Probability Program would need to be made before this is possible. These limitations would lead to good opportunities for future work.

References

[23a] Nov. 2023. URL: https://en.wikipedia.org/wiki/List_of_Wimbledon_gentlemen%27s_singles_champions.

[23b] Dec. 2023. URL: https://www.atptour.com/en/news/best-of-2023-rivalries-djokovic-alcaraz.

[ATP] URL: https://www.atptour.com/en/stats/stats-home.

[Hal21] David Hale. *Is momentum real? an in-depth investigation of sports' most overused term*. Dec. 2021. URL: https://www.espn.com/college-football/story/_/id/32910904/is-momentum-real-depth-investigation-sports-most-overused-term.

[Nag22] Utathya Nag. *Tennis courts: The dimensions, types and all you need to know*. June 2022. URL: https://olympics.com/en/news/tennis-court-markings-dimensions-size-types-variety-surface-hard-grass-clay.

[Ott22] Laura Otto. *The fans are right: When it comes to football, momentum is real*. Nov. 2022. URL: https://uwm.edu/news/the-fans-are-right-when-it-comes-to-football-momentum-is-real/.

[Pre23] Associated Press. *Novak Djokovic to meet Carlos Alcaraz in Dream Wimbledon final*. June 2023. URL: https://www.espn.com/tennis/story/_/id/38010361/novak-djokovic-beats-jannik-sinner-reach-wimbledon-final.

[Rao20] Alex Raoslash;uf. *How do men women serves differ in pro tennis*. July 2020. URL: https://medium.com/@alexrauf/how-do-men-women-serves-differ-in-pro-tennis-ca66ee9e2492#:~:text=For%20first%20serves%20%3A,first%20serve%20points%20than%20Women..

[RBd22] Paul J. Roebber, Bryan M. Burlingame, and Anthony deWinter. “On the existence of momentum in professional football”. In: *PLOS ONE* 17.6 (June 2022). DOI: 10.1371/journal.pone.0269604.

[Sok24] Alexandre Sokolowski. *The day federer and Nadal played "battle of surfaces"*. May 2024. URL: https://www.tennismajors.com/others-news/may-2nd-2007-the-day-that-roger-and-rafa-played-the-battle-of-surfaces-tennis-majors-178677.html.

Appendix A

Wimbledon Data

Filename: wimbledon_serve_win_prob.csv

Wimbledon Data Definitions

Filename: Updated_data_dictionary.csv