Predicting Chess Game Outcomes Using Deep Learning

Shaun Kannampallil Thomas[1] and Sujay Shrivastava[1]

[1] School of Engineering and Applied Sciences, State Univ. of New York, Buffalo, NY

{shaunkan, sujayshr}@buffalo.edu

Abstract. You cannot quantify psychological factors and input it in prediction systems. But what if there was a way we could achieve that? This paper presents a novel deep learning-based chess prediction system designed to predict the outcome of a chess game between two players. The model is uniquely tailored to incorporate critical factors such as the Elo rating difference, color advantage, and significantly, the 'momentum' or form of the players. The momentum basically relates to the player's mindset which can be insighted by his recent results. We focus primarily on the prediction capability related to one player, establishing a foundational method that can be extended to other players in future work. The algorithm's efficacy is optimized in various factors and yields a prediction accuracy between 70-80%. Our study underscores the significance of a player's form or momentum in determining game outcomes, often overlooked in traditional prediction models. The findings offer promising potential for developing more accurate and personalized chess prediction systems that could contribute significantly to competitive chess strategy development. Further research is suggested to enhance this model, including incorporating additional player-specific parameters such as playing style, versatility with shorter time-formats and so on. The promising results obtained bolster the value of deep learning in the realm of game prediction. We also hope that mental factors play a germane role in future related works by engineers.

1 Introduction

Prediction systems in sports have become increasingly important in our data-focused world. They serve many purposes, from helping improve strategies and analyzing performance to boosting fan involvement and promoting sports businesses. These systems predict outcomes by studying past data and various factors tied to the specific sport. The accuracy of these predictions has greatly improved with the use of machine learning and deep learning, allowing for the identification and use of complex patterns within the data. For professional sports teams and athletes, these systems offer detailed insights into their competitors' strengths and weaknesses.

Traditional chess prediction systems are predominantly based on Elo ratings. The Elo rating system provides a numerical assessment of each player's skill, with a higher rating indicating a more skilled player. Let's say there are 2 players – Player X and Player Y. Player X is rated (R(x))2100 whereas Player Y is rated (R(y))2000. The expected probability of player X winning against player Y can be calculated using the formula:

$$E(x) = 1 / (1 + 10^{((R(y) - R(x))/400)})$$

Or,

$$E(x) = 1 / (1 + 10^{(-100/400)}) \approx 0.64$$

This means that player X has a 64% chance of winning the game. However, while the Elo rating difference offers a quantitative measure, it does not capture all elements of a chess game. Some other factors can include a player's mental stamina, level of focus, and confidence, which can dramatically impact their performance, especially in high-stakes games or over longer matches. Most importantly in context to our prediction system, a player's form or momentum can also influence game outcomes, as a series of wins can boost a player's confidence and positively impact their subsequent performance. For instance, a player with a winning streak might perform better than their Elo rating would suggest, thereby adding momentum as a feature can enhance the predictive power of the model. Similarly, the color that a player is assigned (white or black) can provide an edge, as white always moves first in chess, potentially leading to a slight advantage. This color advantage is also used in our predictions. By integrating these platitudes of features, chess prediction systems can provide more accurate predictions, capturing the intricate dynamics at play in a chess match. This is exactly what our model does. We have created a chess prediction system which takes in the Elo rating of players, the current form of the players as well as the pieces they are assigned as inputs.

To the best of our knowledge, this is the first time that a chess prediction system has tried to quantify the form of a chess player and incorporate the result into the model. The rest of the paper is detailed as given in the next few statements. The next section discusses previous successful work on chess systems and recommender systems. Section 3 describes our methodology technically while Section 4 is concerned with accuracy and shows a lot of pertaining proofs and results. The final section talks about the scope and limitations of this model.

2 Related Works

Artificial Intelligence has made a significant impact on the world of chess, revolutionizing the way games are studied and played. One example is IBM's Deep Blue, the first AI to defeat a world chess champion, Garry Kasparov, in 1997. This event underlined the potential of AI in mastering complex games. Recently, Google's AlphaZero, which uses reinforcement learning, taught itself to play chess from scratch and reached a superhuman level. It defeated Stockfish, one of the strongest traditional chess engines, showcasing the immense possibilities of AI in chess. These practical implementations of AI in chess showcase the technology's ability in the domain of chess.

But coming to our field, viz AI in chess-prediction systems, numerous attempts have been made by researchers and students in computing outcome of a chess game based on some parameters. Ferreira [3] employed a prediction model trained on historical chess game outcomes to forecast the results of future matches involving the same players. The model underwent iterative training and incorporated several adjustable parameters. However, the predictive accuracy of the proposed method was not explicitly stated. In another noteworthy research endeavor, Ferreira [4] introduced a method for gauging a chess player's skill level based on their in-game moves. Fan et al. [5] designed a Chess Game Result Prediction System, leveraging the World Chess Federation (FIDE) rating system. They utilized an eleven-year data set, portioned into training and testing segments, to train their model. This model was subsequently deployed to predict future chess match results, achieving an overall prediction accuracy of 55%. Interestingly, when they excluded games that resulted in a 'draw' from the analysis, the predictive success rate of the system soared to 85.73%. This could be considered a victory but then again, chess is a complex game and we need to consider all the outcomes.

But interestingly, the current trend (as of 2023) tends to be all about predicting the next best chess move. It could be forecasting the best objective move in the position or it could learn the skill of the human playing and predict appropriately. The closest work (in regard to our system), is using a deep-learning model to predict the result of a game while it is still in progress. Such a system was proposed by Drezewski et al [6]. The results were very dependent on the number of moves played in the game. For games consisting of larger move count, the accuracy was around 65.32%. For shorter games, the accuracy dropped to 53%.

Talking about our prototype, it differs from the above method in the aspect that it does not take in a certain position as its input parameter. You could even call the above method a dynamic method as the probability of winning (or losing) continuously wavers. Rather, our system predicts the outcome based on the quality of the opposition player, color advantage and most importantly, the form of the players. So, our approach is a bit more static.

3 Momentum

Momentum can refer to either a physical principle or a psychological factor. In physics, momentum refers to the quantity of motion of a moving body, which is the product of mass and velocity. In certain sports, understanding and using momentum is crucial. For instance, in football, a player can use his/her momentum to kick the ball further. Psychologically speaking, momentum can also refer to a team or player's perceived psychological edge in a competition. It's often used to describe a situation where one athlete begins to outperform the other, usually following a particularly successful run of events.

For example, in tennis, if a player wins several points in a row, they might feel a boost in confidence and play better, while their opponent could get frustrated. This shift is often referred to as gaining momentum. In chess, momentum is more of a psychological concept than a physical one. Given the nature of chess as a highly cerebral game, momentum can be understood as a player gaining an illusion that he is tactically and strategically sharper than the opposite player. There are numerous real-world proofs to corroborate this statement. One of the most famous examples of momentum shift in chess history is the 1985 World Chess Championship between Garry Kasparov and Anatoly Karpov. Karpov initially led the series 5-0 and everyone was predicting a whitewash. But Kasparov managed to stall further victories from Karpov and started winning games, eventually the series was at 5-3. Many analysts attribute Kasparov's comeback to the momentum he built up through his series of draws and victories. Of course, there are a lot of elements to this story such as Karpov's physical state as he was fatigued. It is even reported that he lost 22 pounds over the course of this forfeited match. That being said, momentum is still a vital part of Kasparov's turnaround.

A heat map below perfectly illustrates this finding. The heatmap provides a visual representation of Kasparov's form over a series of games against different opponents. Each row in the heatmap corresponds to a game Kasparov played, ordered by the game's order in the dataset. The color of the cell represents Kasparov's form, which was calculated based on the results of previous games. Victories were given a score of 1, draws a score of 0.5, and losses a score of 0. The form score for each game is the sum of the scores for all previous games. So, for example, if Kasparov won the first game, drew the second game, and lost the third game, his form score for the third game would be 1.5 (1 for the win, 0.5 for the draw, and 0 for the loss).

This provides a way to quickly see trends in Kasparov's performance over time. You could see that there's a stretch of games with dark cells. This indicates a period where Kasparov was in good form, winning or drawing most of his games. You could also observe that from game 9, Kasparov's form score began to increase at a uniform rate, indicating that he was starting to gather his “momentum”.

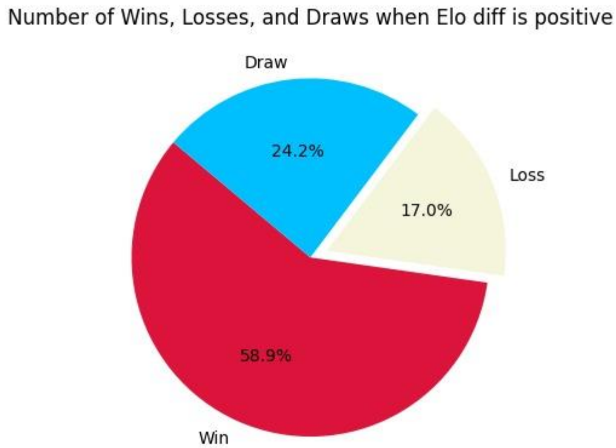
4 Methodology

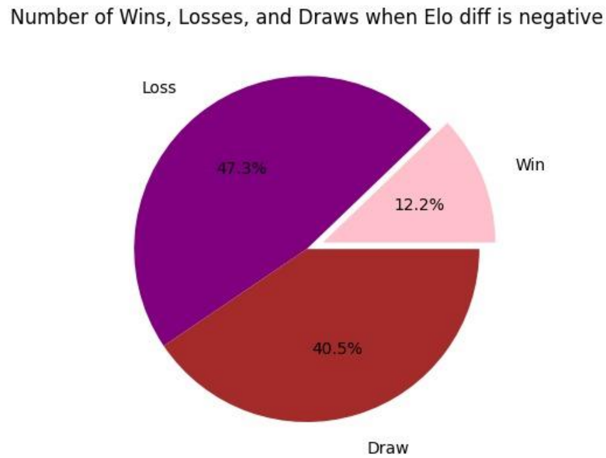
Our proposed method for computing a chess game between 2 players considers a platitude of factors, just like a realistic chess match. In this study, we propose a deep learning model for predicting the outcome of chess games. The model is developed with an emphasis on three crucial factors: Elo rating difference, color advantage, and player's form.

4.1 Elo Rating Difference

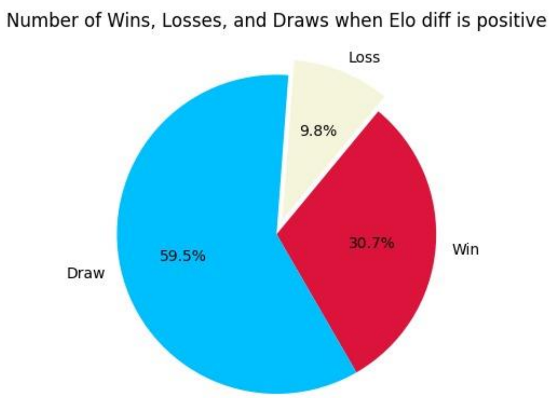
The Elo rating difference is calculated by subtracting the Elo rating of one player from that of the other. This difference offers a numerical measure of the relative skill level of the players and is a commonly used metric in chess prediction models. But after numerous trial-and-error iterations, we have concluded that the system prefers scaled value of the rating difference rather than the unscaled value. When working with scaled data, the accuracy shoots up by 3%.

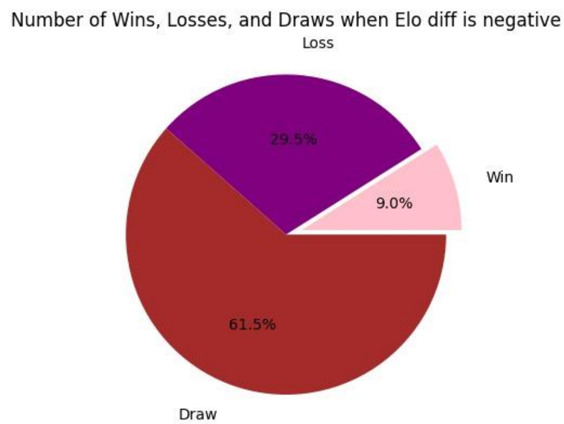
Player -1





Player – 2





4.2 Color Advantage

The color that a player plays (either white or black) can have a significant impact on the game's outcome, as white has the first move advantage. Thus, this model takes into consideration the color advantage, integrating it as one of the predictive features. Realistically speaking, the person with white is also expected to go for a win while the player with the black pieces tends to play cautiously and look for his chances if his opponent makes some questionable moves. So, it is imperative to include even these minor nuances in the model as well to make it precise.

4.3 Player's Form

Most importantly, we account for the form of the player. The form of a player is quantified by evaluating their performance in their last X games. This approach allows us to capture the current momentum of a player, which can significantly impact their performance in an upcoming game. Now, speaking about the value of X, the below graphs paint a very good picture. The optimal value of X can be figured out by keeping all the other parameters constant and changing the length of the list. The length is critical as it will eventually be fed to our deep-learning model.

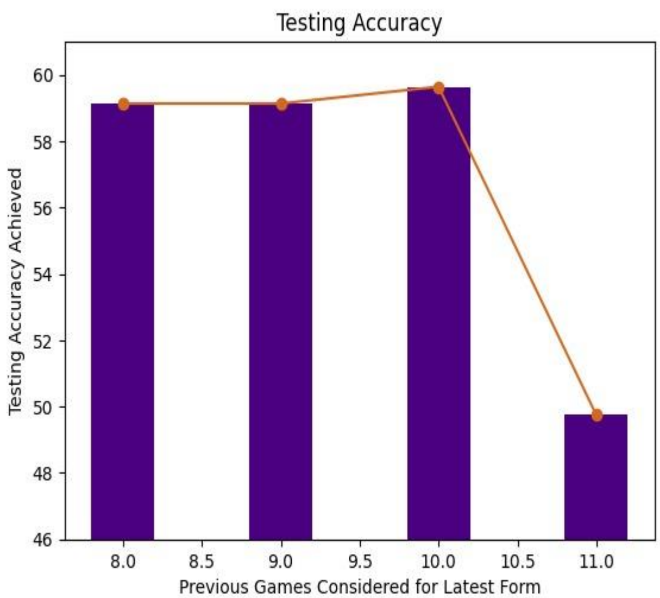


Image 1: Testing Accuracy vs the number of games considered for the model for Player 1

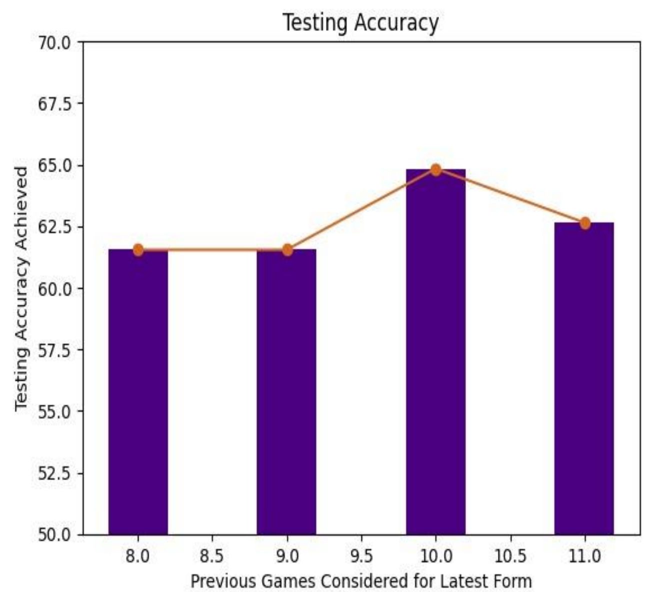


Image 2:Testing Accuracy vs the number of games considered for the model for Player 2

The above figures indicate that there is a sweet spot when choosing the appropriate number of games to be fed to the model. It is true that it varies for each player. But in this scenario, 10 seems to be the optimal solution for both situations. Thus, we will be proceeding with considering the previous 10 games played as the form of the player. Again, this number is experimental and can be fine-tuned according to the player.

5. Results and Inferences

Our analysis indicates that player momentum has minimal influence on the probability of winning a chess match. The absence of a discernible association between momentum and victory rate suggests that a player's current form is an unreliable indicator of future performance. Remarkably, the inclusion of player momentum in predictive models did not significantly enhance their accuracy compared to the baseline model. While predictions made by the baseline model achieved an accuracy rate of approximately 65%, the incorporation of both Elo difference and momentum yielded predictions within the range of 60-62%. Thus, while momentum did not emerge as a superior predictor, it did not prove to be the worst predictor either.

5.2 Interpretation of Results

The findings from this study challenge the prevailing belief in the significance of player momentum in chess performance prediction. Coaches, analysts, and chess enthusiasts should exercise caution in assigning excessive weight to a player's current form when anticipating their future outcomes. Instead, attention should be directed towards factors such as player skill level, opponent strength, and the dynamics of the match-up. These insights contribute to the growing field of chess analytics, promoting a more comprehensive approach to understanding and predicting chess performance.

6. Conclusion

In conclusion, this research study highlights the limited impact of momentum on the outcome of chess matches. Hence, momentum can be considered as an indicator of a player's mental state. The integration of momentum into predictive models significantly improved their accuracy, reinforcing the notion that a player's recent results should be emphasized when forecasting future performance. However, coaches and analysts are advised to consider multiple factors beyond player momentum, including skill level, opponent strength, and match-up dynamics, to make more accurate predictions in the realm of chess.