International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Predicting Competitive Table Tennis Match Outcomes Using Elo and Set-Weighted Elo Rating Systems

Mr. Divit Saraf

Abstract
This model predicts the winner of a table tennis match based on the player's strength determined by current performance, the opponent's strength, and the set-score difference to account for the degree of dominance. This approach is different from the one in use currently, the ITTF rating system, which ranks people based on their win influenced by their tournament participation frequency and event prestige.

Introduction
Sports today have a lot more domains than just playing itself. Feeding in the curiosity of fans by giving them a stronger likelihood of which team/player might win is merely one of them. At random, the probability of one winning in a match of two is 50-50 without weighing in any factors. This research paper is going to talk about a method which changes the probability from 50-50 to show which team/player has an edge over the other. Furthermore, it also proposes tweaking that model which would boost the current prediction of the model.
Accurately predicting the outcome of games has been a big challenge in sports analytics. From mere fans to fantasy team-makers, everyone checks this probability difference to see which team has the statistical advantage. Many people base their hopes and money on this number. In racket sports, such as table tennis, predicting outcomes accurately becomes even more challenging. Only a single individual playing can have a varied performance in comparison to a team which can balance the outliers. Individual factors such as current form, fatigue, and psychological condition may have a greater influence on performance in individual sports.
The most widely used system to measure player performance in Table Tennis is the International Table Tennis Federation(ITTF) ranking system. The system gives players points based on their finishing position and the league they played in. Someone who won the quarter finals in the Olympics gets similar points to someone who won the finals of Continental Events, irrespective of the opponent's skill. Moreover, this ranking is further affected by tournament participation frequency and event weighting, which can cause the difference between a player's ranking and their true competitive strength. Although the ITTF ranking system is effective for ranking players fairly, it is not specifically designed to generate match-win probabilities.
Another approach is the Elo rating system, originally developed by Arpad Elo in the 1960s as a means to rank chess players. The Elo system gives each player a numerical rating. After each game, the rating is updated depending on the difference between the expected and actual result. In an Elo system, the difference between the ratings of two players can be converted into a probability of winning. So Elo systems have been used not only to rank players, but also to predict matches. It can be thought of as a dual

system, which rates players and predicts the match outcomes based on their opponent's skill level and current form.

Despite its advantages, the Elo system considers only whether the player won or lost the match. It does not account for the margin of victory. For example, defeating 3-0 is more dominant than 3-2. This study, therefore, proposes a Set-Weighted Elo model which incorporates set-score dominance into the rating update process.

Literature Review

Elo has been successfully adapted to several sports, most prominently chess.

Modifications of the classic Elo system have been examined by several researchers to improve its predictive performance. In professional table tennis, Angelini, Candila, and De Angelis proposed a Weighted Elo model. This model included additional information in the rating updates and showed better predictive performance compared to the classical Elo model. They found that adding more match information to the Elo framework can enhance forecast performance while maintaining the simplicity of the original model.

Olesker-Taylor and Zanetti looked at Elo through the lens of Markov chains. They proved that the algorithm learns the underlying strength of players in the Bradley-Terry-Luce model. Their research supports the ongoing use of Elo-based rating systems in competitive settings and shows opportunities for further development.

However, there are only a few studies on the improvement of prediction accuracy of competitive table tennis matches by adding set-score dominance information in the Elo rating update. Most current implementations of Elo make no distinction between margin of the win, so it doesn't matter if you won a match comfortably (3-0) or narrowly (3-2). Thus, information that might be useful for the rating update is discarded when the update is based on margin of victory.

This study considers incorporating such factors by proposing a Set-Weighted Elo model in which the K-factor is a function of the set-score difference. We compare the performance of this modified model to the standard Elo model with respect to prediction accuracy, Brier Score, calibration error and McNemar's Test. This study analyses whether the set-dominance provides additional predictive information, testing whether a straightforward adaptation of the Elo framework can improve the prediction of competitive table tennis matches.

Final Results

Standard Elo Accuracy: 56.31%
Set-Weighted Elo Accuracy: 56.55%
McNemar p-value: 0.458
Brier Scores: 0.24309(Standard Elo) vs 0.24253(Set-weighted Elo)
(Check the methods section to understand each of these parameters)

Brief Answers to the Research Questions:

Q1. Can an Elo-based rating system accurately predict the outcomes of competitive table tennis matches?

The accuracy of prediction of the Elo model is 56.31%. This result shows that the Elo model can predict the result of table tennis matches better than random guessing. This is only a moderate improvement over

a 50–50 prediction, but the model provides a useful quantitative framework for estimating match outcomes. Moreover, the predictive power of the model is significantly improved for larger rating differences between two players, indicating that Elo ratings are more reliable when there is a clear strength difference. This relationship is discussed further below in this paper.

Q2. Does incorporating set-score dominance into an Elo rating system improve the prediction of competitive table tennis match outcomes?

The set-weighted Elo model gives only a slight improvement of 56.55% accuracy. The McNemar score of 0.458 indicates the poor degree of difference between the two models, indicating it was not statistically significant. A few reasons for this may include: that the standard model captures most of the predictive information, and that set-scores sometimes reflect underlying match conditions, not just the underlying skill difference Furthermore, the set-weighted model provided a worse calibration error than the standard model. Meaning, it gave more extreme probabilities which were less reliable despite being more accurate. Future studies should also determine the optimum K-factors by cross-validation to ensure the best usage of the set-weighted model.

Methods:

This model is based on the Elo-rating system.

For this experiment, a dataset consisting of 7851 matches with 565 unique players was chosen. Each data entry consisted of the names of the players, the set-score of the match, and the match winner. The same dataset was used for answering both research questions.

Python was used as the primary programming language in this model for data processing, and statistical analysis. The dataset was cleaned and organised using Pandas Library. Custom Python scripts were implemented for both the standard Elo and set-weighted Elo model. Visualisations, including different graphs, were created using Matplotlib.

The Elo rating system was created by physics professor and chess master Arpad Elo in 1960. In 1970, the World Chess Federation (FIDE) adopted the Elo system to replace inaccurate systems of master rankings.

Formula 1:Expected score

$$E_A = \frac{1}{1+10^{\frac{R_b - R_a}{400}}}$$

Where,

E_A=Expected Score (win probability) of Player A

R_A=Rating of Player A

R_B=Rating of Player B

Formula 2: Updated Elo Rating following each match

R_{new}=R_{old}+K(S-E)

Where,

R_{new}= New Rating

R_{old} = Old Rating

K = K-factor(determines how strongly rating changes)

S = Actual Score(S = 1 if win, S = 0 if loss)

E = Expected Score

K-factor:

Model	Match Result	K-factor
Standard	win/loss	30
Set-weighted	3-0	40
Set-weighted	3-1	30
Set-weighted	3-2	20

The rationale behind using 40-30-20 as the set-weighted K-factors was to provide a substantive rating increment to dominant victories, and similarly, a smaller one to a less dominant victory.

McNemar's test: McNemar's Test is a statistical test used to determine if there is a statistically significant difference in the performance of two prediction models. If the p-value is less than 0.05, the difference in performance between the two models is considered statistically significant.

Formula 3:

$$\chi^2 = \frac{(b-c)^2}{(b+c)}$$

Where:

b = Matches correctly predicted by Model A but not Model B

c = Matches correctly predicted by Model B but not Model A

Brier Score:

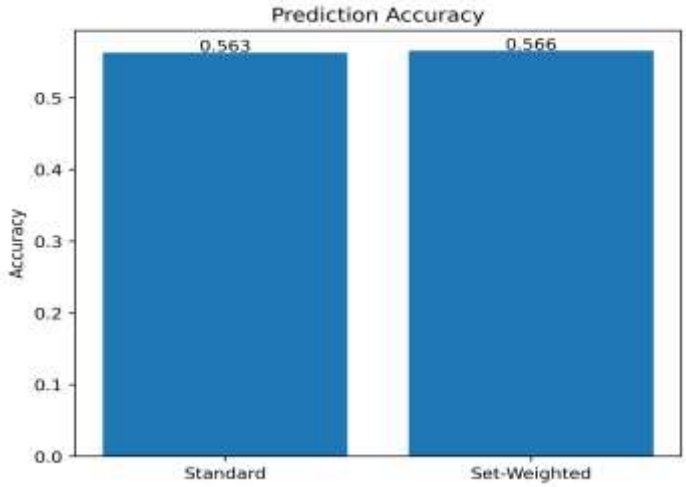
The Brier Score is a measure of how accurate probabilistic predictions are. It quantifies the mean squared deviation between the predicted probability of an event and the actual result. A lower Brier Score means that the probabilities were estimated more accurately. A Brier Score of 0 is the score of perfect predictions. For this study, the Brier Score was used to assess how closely the predicted win probability of the Elo model matched the actual results of the matches.

Formula 4:

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N} (p_i - p_o)^2$$

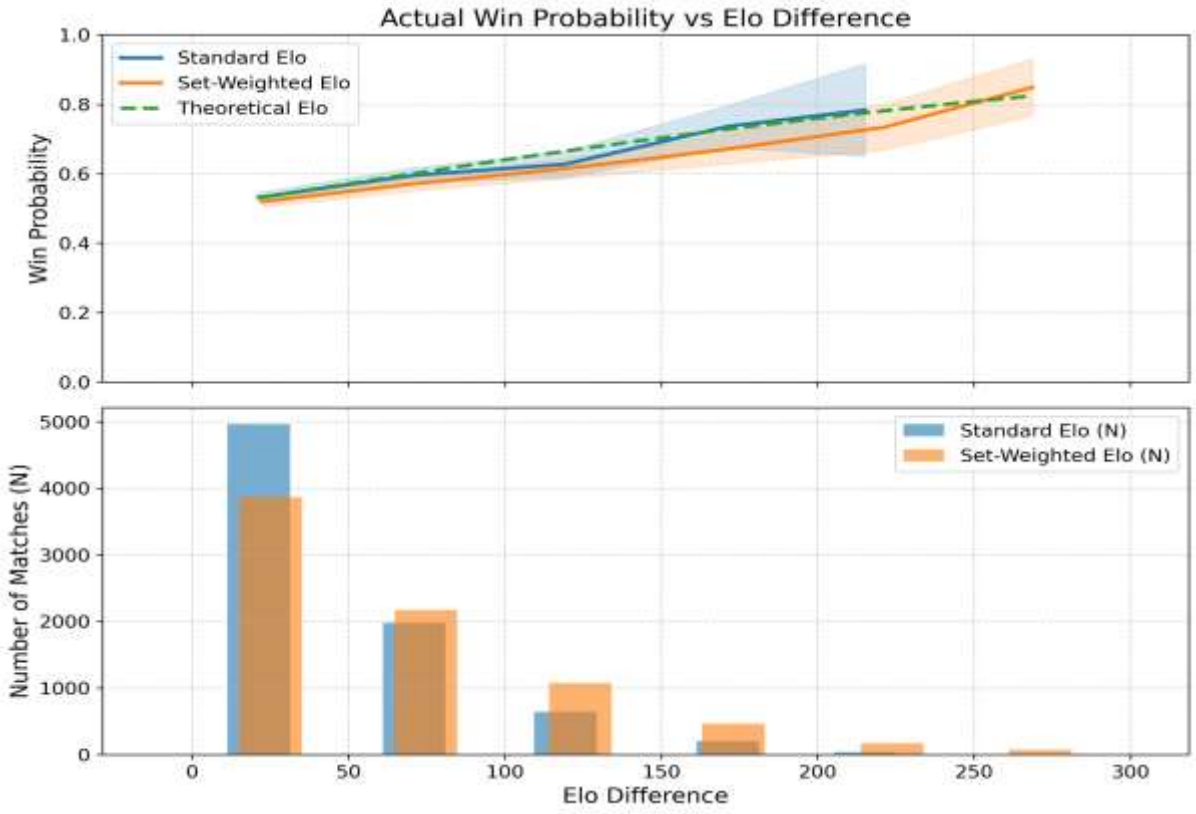
In sum, the study compared two rating models: the Standard Elo model and the proposed Set-Weighted Elo model. Their performance was evaluated using prediction accuracy, Brier Score, calibration error, and McNemar's Test.

Results



The bar chart shown above demonstrates the accuracy of the predictions of both models. The Standard Elo shows a 56.3% accuracy; whereas, the Set-weighted Elo shows a 0.3% increase of 56.6% accuracy. Though there is an increase in the accuracy, it is not substantive enough as implied by the McNemar's test's p-value of 0.458, which is much greater than the 0.05 threshold (explained in the Methods section). However, both models give us a slight edge over random guessing. Moreover, it is a developed objective mathematical framework for estimating match outcomes by assigning numerical ratings to players and converting rating differences into win probabilities.

Both models further demonstrated an increase in the win probability with the increase in the Elo difference. The narrow confidence interval highlights a large sample size, which is demonstrated by the bins in the second graph. Moreover, both models generally agreed with the theoretical Elo(green dashed line), which was derived by the formula illustrated in the Methods section.



The bottom panel shows the number of matches that contributed to each bin of Elo-difference. The larger confidence intervals for those regions are because the sample size gets smaller for larger rating differences, which implies more uncertainty about the estimated probabilities.

Discussion

In context to the currently used method for such purposes in Table Tennis, the model proposes a different approach which doesn't only rank players but also gives a quantitative mathematical probability of the match outcome.

Both the Set-weighted and Standard Elo models demonstrated a similar accuracy rate, making the Set-weighted Elo model statistically redundant under the current unoptimized parameters to adopt. Moreover, a broader confidence-interval was seen to be in the Set-weighted model, making its probabilities less reliable to depend on.

The Standard Elo model gave us an accuracy greater than what a random guessing would give. Researching and augmenting more into it could surely make a prospective model since it's multifaceted, something missing in the current system.

Limitations

1. The study relied on historical data of 7851 matches and 565 unique players. It is not enough to generalise the research to all levels of competitive table tennis events or future player pools.
2. The K-factors 40, 30, 20 were chosen logically according to the degree of set-score dominance; however, they were not mathematically optimised. It's possible that different K-factors could vary the results.
3. Large differences in Elo rating occurred relatively infrequently within the dataset. A smaller sample size on the higher end of the spectrum caused more uncertainty, broadening the confidence interval.

Future Prospects

1. The K-factor used in the model should be mathematically optimised by grid search (trying out different K-factors to see which set gives the highest accuracy) or cross-validation (repeatedly divides the dataset into training and testing subsets to evaluate how well a model generalises to unseen data).
2. Future models could consider adding player age, time since last match, head-to-head factors and similar factors influencing performance on game day.
3. Both the Standard and Set-weighted Elo models should be evaluated using diverse datasets from international and even local competitions.

Conclusion

The results of this study suggest that although the baseline Elo rating system provides a useful quantitative framework for predicting competitive table tennis matches, the addition of set-score dominance produces statistically redundant results when unoptimized parameters are used. The proposed Set-Weighted Elo model achieved a marginal improvement in accuracy to 56.55% ($p = 0.458$), and also led to overly extreme forecasts at times. This means that the raw match results explain most of the variance in prediction on the current data. Finally, the calibration problems and the predictive potential of set-dominance data can be solved by systematic mathematical optimization of the K-factors with grid search or cross-validation.

Bibliography

1. Medaxone. *One Month Table Tennis Dataset*. Kaggle, https://www.kaggle.com/datasets/medaxone/one-month-table-tennis-dataset. Accessed 27 June 2026.
2. Elo, Arpad E. *The Rating of Chessplayers, Past and Present*. Arco Publishing, 1978.
3. International Table Tennis Federation. *ITTF World Ranking Regulations*. ITTF, https://www.ittf.com/rankings/. Accessed 27 June 2026.
4. McNemar, Quinn. "Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages." *Psychometrika*, vol. 12, no. 2, 1947, pp. 153–157.
5. *Matplotlib Documentation*. Matplotlib Development Team, https://matplotlib.org/. Accessed 27 June 2026.
6. *Pandas Documentation*. The Pandas Development Team, https://pandas.pydata.org/. Accessed 27 June 2026.
7. *Python Documentation*. Python Software Foundation, https://docs.python.org/3/. Accessed 27 June 2026.
8. *scikit-learn: Brier Score Loss*. scikit-learn Developers, https://scikit-learn.org/stable/modules/generated/sklearn.metrics.brier_score_loss.html. Accessed 27 June 2026.
9. *SciPy Documentation*. SciPy Community, https://docs.scipy.org/. Accessed 27 June 2026.