Revista de Artes Marciales Asiáticas
Volume 19(2), 171-181 ~ July-December 2024
DOI: 10.18002/rama.v19i1.2414
http://revpubli.unileon.es/ojs/index.php/artesmarciales

RAMA
I.S.S.N. 2174-0747

Forecasting judo medal winners at the Olympic Games: an interaction of the International Judo Federation World Ranking List and the Elo System

Leandro Marques GUILHEIRO[1] & Emerson FRANCHINI[*1]

[1] *Institute for Pure and Applied Mathematics (IMPA) and Esporte Clube Pinheiros (Brazil)*

[2] *Martial Arts and Combat Sports Research Group, Sport Department, School of Physical Education and Sport, University of São Paulo (Brazil)*

Received: 06/05/2024; Accepted: 13/11/2024; Published: 14/11/2024

BY NC SA

ORIGINAL PAPER

Abstract

The aim of this study was to verify which ranking list predicts Olympic results better: one created by Elo system, one using the International Judo Federation (IJF) World Ranking List (WRL), or another using the combination of both. The data utilized comprised the outcomes of 93,728 matches, encompassing 42,844 matches from the 2016 Rio Olympic Games cycle and 50,884 matches from the 2020 Tokyo Olympic Games cycle. These matches were held across 311 events, all of which contribute points to the IJF WRL. The data was sourced from https://judobase.ijf.org. A total of 8,142 male and 4,736 female judo athletes from all weight categories were analyzed. We employed two variables as proxies for athletes' performance throughout the Olympic cycle: the positions in the IJF WRL and the ratings from the Elo System. A binary-response model was utilized. In this model, "success" denoted an athlete receiving a medal, while "failure" indicated otherwise. A combination of the WRL and Elo system better predicted Olympic performance of judo athletes. Additionally, for each rank position an athlete improved in the IJF WRL, there was an increased probability to win an Olympic medal of approximately 7.50%, while for each 10 Elo rating score improvement, the athlete increased the probability to win an Olympic medal in approximately 9.26%. When both systems were used together, the accuracy of the model was approximately 91%, with a sensitivity of nearly 68-69%, and a specificity close to 95%, for Rio de Janeiro and Tokyo editions isolated or grouped. Such information can serve as a valuable tool for national federations staff in selecting the most suitable athletes to participate in the Olympic judo competition, if both the WRL and an Elo rating system are used together.

Keywords: Martial arts; combat sports; judo; performance prediction; Olympics; rating.

Predicción de medallistas olímpicos en judo: una interacción entre la Lista de Clasificación Mundial de la Federación Internacional de Judo y el Sistema Elo

Resumen

El objetivo de este estudio fue verificar qué lista de clasificación predice mejor los resultados olímpicos: una creada por el sistema Elo, una que utilice la Lista de Clasificación Mundial (WRL) de la Federación Internacional de Judo (IJF), o una combinación de ambas. Los datos utilizados fueron los resultados de 93728 combates, que incluyeron 42844 combates del ciclo de los Juegos Olímpicos de Río 2016 y 50884 del ciclo de los Juegos Olímpicos de Tokio 2020, desarrollados en 311 eventos que sumaron puntos al WRL de la IJF. Los datos se obtuvieron de https://judobase.ijf.org. Se analizaron 8142 atletas masculinos y 4736 femeninos de judo de todas las

Previsão de vencedores de medalhas de judô nos Jogos Olímpicos: uma interação entre a Lista de Classificação Mundial da Federação Internacional de Judô e o Sistema Elo

Resumo

O objetivo deste estudo foi verificar qual lista de classificação prevê melhor os resultados olímpicos: uma criada pelo sistema Elo, uma utilizando o Ranking Mundial (WRL) da Federação Internacional de Judô (IJF), ou outra utilizando a combinação de ambas. Os dados utilizados compreenderam os resultados de 93.728 combates, abrangendo 42.844 combates do ciclo dos Jogos Olímpicos do Rio 2016 e 50.884 combates do ciclo dos Jogos Olímpicos de Tóquio 2020. Esses combates foram realizados em 311 eventos, todos os quais contribuem com pontos para o WRL da IJF. Os dados foram obtidos em https://judobase.ijf.org. Um total de 8.142 atletas do sexo masculino e 4.736 atletas do sexo feminino de judô de todas as categorias de peso

* *Corresponding author*: Emerson Franchini (efranchini@usp.br)

Contributions: Leandro Marques Guilheiro (ABCEFGIKMN), Emerson Franchini (AEFGHJKMN). Codes according to CRediT (Contributor Roles Taxonomy): (A) Conceptualization. (B) Data curation. (C) Formal Analysis. (D) Funding acquisition. (E) Investigation. (F) Methodology. (G) Project administration. (H) Resources. (I) Software. (J) Supervision. (K) Validation. (L) Visualization. (M) Writing – original draft. (N) Writing – review & editing.

Funding: The authors received no funding for this work.

Conflicts of interest: The authors report there are no competing interests to declare.

© 2024 by the authors. Licensee Universidad de León, Spain. www.unileon.es

categorías de peso. Empleamos dos variables como indicadores del rendimiento de los atletas a lo largo del ciclo olímpico: las posiciones en el WRL de la IJF y las puntuaciones del Sistema Elo. Se utilizó un modelo de respuesta binaria. En él, "éxito" denotaba a un atleta medallista, y "fracaso" indicaba lo contrario. Una combinación del WRL y el sistema Elo predijo mejor el rendimiento de los atletas. Además, por cada posición que un atleta mejoraba en el WRL de la IJF, aumentaba aproximadamente un 7.50% la probabilidad de que este ganase una medalla olímpica, mientras que por cada mejora de 10 puntos en la puntuación Elo, el atleta aumentaba esta probabilidad en aproximadamente un 9.26%. Al utilizar juntos ambos sistemas, la precisión del modelo fue aproximadamente del 91%, con una sensibilidad cercana al 68-69% y una especificidad cercana al 95%, para las ediciones de Río de Janeiro y Tokio, ya sea aisladas o agrupadas. Esta información puede ser útil para las federaciones nacionales, para seleccionar a los atletas más adecuados para participar en la competición olímpica de judo, si se utilizan en conjunto el WRL y un sistema de puntuación Elo.

Palabras clave: Artes marciales; deportes de combate; judo; predicción de rendimiento; Juegos Olímpicos; clasificación.

foram analisados. Foram empregadas duas variáveis como proxies para o desempenho dos atletas ao longo do ciclo olímpico: as posições no WRL da IJF e as classificações do sistema Elo. Um modelo de resposta binária foi utilizado. Neste modelo, "sucesso" denotava um atleta recebendo uma medalha, enquanto "fracasso" indicava o contrário. Uma combinação do WRL e do sistema Elo previu melhor o desempenho olímpico dos atletas de judô. Além disso, para cada posição de classificação que um atleta melhorou no WRL da IJF, houve uma probabilidade aumentada de ganhar uma medalha olímpica de aproximadamente 7,50%, enquanto para cada melhoria de 10 pontos na pontuação Elo, o atleta aumentou a probabilidade de ganhar uma medalha olímpica em aproximadamente 9,26%. Quando ambos os sistemas foram usados juntos, a precisão do modelo foi de aproximadamente 91%, com uma sensibilidade de cerca de 68-69% e uma especificidade próxima a 95%, para as edições do Rio de Janeiro e Tóquio isoladas ou agrupadas. Essas informações podem servir como uma ferramenta valiosa para as federações nacionais na seleção dos atletas mais adequados para participar da competição olímpica de judô, se tanto o WRL quanto um sistema de classificação Elo forem utilizados juntos.

Palavras-chave: Artes marciais; desportos de combate; judo; previsão de desempenho; Jogos Olímpicos; classificação.

1. Introduction

Since 2009, the International Judo Federation (IJF) has been using a World Ranking List (WRL) to classify judo athletes. This ranking is primarily utilized to determine the athletes qualified to take part in the Olympic Games and to assign their position in the seed during the official IJF competitions (IJF, 2024). Therefore, nations interested in having their athletes competing at the Olympic Games need to carefully choose the competitions in which these athletes participate to maximize the number of points earned while minimizing travel costs and athlete's exposure to injury during this process (Franchini et al., 2017). After the London Olympic Games in 2012, the first edition in which the IJF WRL was used to determine the athletes qualified, several studies were conducted to verify its predictive value for winning a medal during the Olympic Games (Brunel, 2022; Daniel & Daniel, 2013; Franchini & Julio, 2015; Santos et al., 2023). A similar approach was also conducted to verify the WRL's predictive value for winning a medal in junior and senior Judo World Championship (Breviglieri et al., 2018). Additionally, the relationship between the WRL position, the seeding or lack thereof of a given athlete, and the probability of winning a medal was also studied (Brunel, 2022; Guilheiro & Franchini, 2017).

In general, when using the multiple linear regression approaches, studies indicated a moderate to low predictive value of the IJF WRL regarding competitive success in top-level judo competitions (Breviglieri et al., 2018; Franchini & Julio, 2015; Guilheiro & Franchini, 2017; Santos et al., 2023). Conversely, the use of Bayesian (Guilheiro & Franchini, 2017) or Monte-Carlo simulation approaches (Brunel et al., 2022), indicated that being seeded was associated with a higher probability of winning a medal at the judo Olympic tournament. Therefore, predicting the outcomes of judo competitions at the Olympic Games can yield valuable insights for managers, coaches, staff, and athletes. An important aspect to be considered is the difference between ranking and rating the judo athletes, a problem that has affected other sports (Minton, 2017). A rating system generates scores to each athlete based on some rules. When these scores are sorted, a ranking list is created (Langville & Meyer, 2012). Specifically in judo, some athletes deliberately compete less than others (e.g., Shohei Ono, Teddy Riner, Hifume Abe, Uta Abe, etc.), resulting in lower IJF WRL position than expected, but with a very low defeat record. Therefore, if a rating was used to create a different ranking list, these athletes would be in a better position than using the IJF official ranking. Thus, this would affect their positioning in the seed and, consequently, confrontations between athletes who could achieve the podium would be postponed to later phases within the competition, increasing the attractiveness of the event. This aspect is especially important considering that the repechage in judo is conducted only for athletes reaching the quarter-finals; therefore, any confrontation between two top athletes before this phase results in one of them being excluded from the competition. The Elo system (Elo, 1978), created to rate chess players, can be considered an approach to avoid this misclassification of athletes competing only a few times during the year. Briefly, the Elo system updates the athlete's rating after every competition based on his/her performance in that competition, with the variation in the rating reflecting both the result and the quality of the adversary (Minton, 2017). Since its introduction, Elo rating system has been used in many contexts, and it is still used for ranking many contests such as online games, and sports (e.g., football) (Pelánek, 2016). In judo the quality of all the opponents in a given competition would be computed. Thus, the primary aim of this study was to compare which ranking list predicts Olympic results better: one created by Elo system, one using the IJF WRL, or another combining both. The main hypothesis was that the Elo rating system or a combination of these systems could be effectively utilized to predict, with reasonable accuracy, which athletes are likely to win a medal in the subsequent Olympic Games edition after the classificatory competitions analyzed. Such information can serve as a valuable tool for national federations staff in selecting the most suitable athletes to participate in the Olympic judo competition.

2. Methods

2.1. Experimental design

This study adopts a descriptive and predictive analysis approach, focusing on the judo competition of two editions of the Olympic Games while considering judo athletes' previous performance in the IJF World Tour competitions.

2.2. Participants

The data utilized in this study comprises the outcomes of 93,728 matches, including 42,844 matches from the 2016 Rio Olympic Games cycle and 50,884 matches from the 2020 Tokyo Olympic Games cycle. These matches were held across 311 events, encompassing Senior World Championships, World Masters, Grand Slams, Grand Prix, Continental Opens, and Continental Championships, all of which contribute points to the IJF WRL. The data was sourced from https://judobase.ijf.org. A total of 8142 male and 4736 female judo athletes from all weight categories were analyzed. In order to test the model, the results of 390 and 391 athletes who competed in Rio and Tokyo, respectively, were used. Therefore, there was no sampling process because all athletes who participated in these competitions were included in the analysis. Informed consent was not required for the present study, as established by the Belmont Report (1979), as the data was fully available online, no experimentation with the participants, and no identification of the athletes were necessary.

2.3. Variables of study

It is reasonable to assume that the quality of athletes serves as a reliable predictor of the outcomes in a given competition. However, relying solely on the results of a single tournament may not adequately represent the athletes' quality or accurately predict future outcomes. Therefore, in this study, we employed two variables as proxies for athletes' performance throughout the Olympic cycle: the positions in the IJF WRL and the ratings from the Elo System.

- International Judo Federation World Ranking List Position

Each athlete can sum up to a maximum of twelve competition results in the IJF WRL. The points attributed to these results depend on the athlete's final position in the event and the type of event. Athletes accumulate five results on this ranking list, along with an additional result either from World Masters or Continental Championships, spanning from 12 to 24 months, and five results plus one from the present time to 12 months. The points attributed in the former period are reduced by 50%, while the latter counts 100%.

Given the rules of the IJF WRL and its selection function, where one must be among the top 17 athletes, excluding those from the same country, to directly qualify for the Olympic Games, and the ease of utilizing such ranking, it seems reasonable to use it as a proxy for athletes' performance

(IJF, 2024). Moreover, since athletes cannot accumulate points unlimitedly in their ranking, it is necessary to replace existing points by achieving better results in tougher competitions. Therefore, the higher one's position in the IJF WRL, the more points are required to advance to the next ranking position (Guilheiro, 2020). This condition provides an important threshold for evaluating the quality of athletes through this ranking list.

The drawback of this variable is that it cannot capture the variance in one's performance. For instance, one athlete may participate in ten competitions to reach a given ranking position, while another may participate in only three to achieve a similar position. The Japanese team serves as a notable example of athletes who rarely compete in the IJF World Tour, making it difficult for them to attain the leading position in the IJF WRL. However, some of them consistently secure gold medals whenever they compete, including in Grand Slams and World Championships, demonstrating their excellence. In this study, the ranking position of each athlete prior to the Olympic Games was used as published in the following days: (a) Rio: August 4, 2016; (b) Tokyo: July, 22, 2021.

- Elo Rating System

"A proper rating system should go a step beyond mere ranking and should provide some estimate of the relative strengths of the competitors, however strength may be defined" (Elo, 1978). In this sense, the rating system created by Arpad E. Elo for chess was adapted to estimate the strength of judo athletes at two different junctures: the 2016 Rio Olympic cycle and the 2020 Tokyo Olympic cycle.

The first implementation of a similar scheme for judo occurred in 2011, spearheaded by Lance Wicks (Wicks, 2011). He primarily compared the positions of athletes in the IJF WRL with their estimated positions in a parallel ranking using the rating system. In this current study, the rating system was incorporated as a covariate to predict the outcomes of the Olympic Games.

In Elo's scheme, the determination of the k-factor is crucial. The k-factor is a parameter used to determine the rate at which a player's rating changes based on the outcome of a match. In this paper, we estimated the optimum k-factor for each weight category in both Olympic cycles using maximum likelihood estimation. 'The fundamental idea is that we can optimize an objective criterion, i.e., find the k that leads to the best possible value for that criterion. The criterion we can look at is the maximum likelihood of winning probabilities" (Neumann & Kulik, 2020). The initial rating was arbitrarily defined as 1500 for each athlete, as proposed by Elo, and the R package "EloRating" (Neumann & Kulik, 2020) was utilized to implement Elo's scheme and optimize the k-factor.

The ratings were calculated from the first competition following the previous Olympic Games until the final competition preceding the anticipated Olympic Games of interest for prediction: (a) Rio: from the Grand Prix Abu Dhabi 2012 until the Grand Slam Tyumen 2016; (b) Tokyo: from the Asian Open Taipei 2016 until the World Championships Seniors Hungary 2021. Given that the interval between the Olympic Games is four years and that the Elo rating system requires exposure of athletes in competitions to estimate an accurate rating, this time frame appeared suitable for practical reasons.

The constant k-factor for each weight category for each Olympic Games edition is presented in Table 1.

Table 1. Rio de Janeiro and Tokyo Olympic Games k-factors for each weight category.

Weight Category Females (kg)	Rio de Janeiro 2016	Tokyo 2020+1	Weight Category Males (kg)	Rio de Janeiro 2016	Tokyo 2020+1
≤ 48	76	77	≤ 60	72	73
≤ 52	85	89	≤ 66	65	63
≤ 57	79	80	≤ 73	65	69
≤ 63	72	71	≤ 81	71	71
≤ 70	67	75	≤ 90	67	67
≤ 78	86	91	≤ 100	72	71
> 78	80	87	> 100	79	85

2.4. Statistical Analyses

Given the study's objective of predicting athlete performance in the upcoming Olympic Games, the utilization of a binary-response model was needed. In this model, "success" denotes an athlete receiving a medal, while "failure" indicates otherwise. All models were implemented in R for Mac(Intel), version 4.3.3, using the Generalized Linear Models function "glm()".

- Approach 1

We employed the logit model (Cox, 1958) for both the Rio and Tokyo Olympics, individually and together. We utilized three linear predictors to assess which combination best predicts medal outcomes in the Olympic Games:

$$\text{Model 1: } log\frac{\pi_j}{1-\pi_j} = \beta_{1,0} + \beta_{1,1} wrl_j$$

$$\text{Model 2: } log\frac{\pi_j}{1-\pi_j} = \beta_{2,0} + \beta_{2,1} elo_j$$

$$\text{Model 3: } log\frac{\pi_j}{1-\pi_j} = \beta_{3,0} + \beta_{3,1} wrl_j + \beta_{3,2} elo_j$$

where, for $j \in$ N: π_j = expected probability of the *j-th* athlete win a medal; *wrl* = IJF WRL position of the *j-th* athlete; elo_j = Elo rating of the *j-th* athlete

After conducting the inferential analysis of the coefficients in each model, it was possible to calculate the odds ratio for the chosen model. Likelihood ratio tests (LRT) were employed to compare the adjusted and maximal models, assessing their adequacy. Subsequently, the adjusted models were compared against each other to determine the best-fitting model for our data using the Akaike Information Criterion (AIC) and, once more, LRT for the nested models. Finally, we assessed the classification quality through several numerical metrics, including accuracy, sensitivity, specificity, Receiver Operating Characteristic (ROC), and Area Under the Curve (AUC).

- Approach 2

We employed the logit model for all weight categories, individually for both the Rio and Tokyo Olympics. We utilized three linear predictors to assess which combination best predicts medal outcomes in the Olympic Games:

$$\text{Model 1: } log\frac{\pi_{ij}}{1-\pi_{ij}} = \beta_{(1,0)i} + \beta_{(1,1)i} wrl_{ij}$$

$$\text{Model 2: } log\frac{\pi_{ij}}{1-\pi_{ij}} = \beta_{(2,0)i} + \beta_{(2,1)i} elo_{ij}$$

$$\text{Model 3: } log\frac{\pi_{ij}}{1-\pi_{ij}} = \beta_{(3,0)i} + \beta_{(3,1)i} wrl_{ij} + \beta_{(3,2)i} elo_{ij}$$

where, for $i, j \in$ N: π_{ij} = expected probability of the *j-th* athlete of the *i-th* weight category win a medal; wrl_{ij} = IJF WRL position of the *j-th* athlete of the *i-th* weight category; elo_{ij} = Elo rating of the *j-th* athlete of the *i-th* weight category.

Thus, we were faced with a classification task aimed at predicting the medal winners in the upcoming Olympic Games. Therefore, it was necessary to evaluate the model's efficacy by comparing the actual outcomes with the predicted ones using a confusion matrix. This matrix enables the assessment of classification quality through several numerical metrics, including accuracy, sensitivity, specificity, ROC, and AUC. While a high accuracy was desirable, the imbalanced nature of the dataset (with more failures than successes, given that only four athletes can win a medal in each weight category) could potentially be misleading. Hence, sensitivity emerges as the primary indicator of our model's predictive capability, as it identifies the predicted medal winners who actually receive medals. Given our interest in identifying the top four athletes in each weight category, the threshold probability for classifying as "success" varied for each category.

3. Results

3.1. Approach 1

Tables 2, 3 and 4 present the main results for each of the three models, respectively.

Table 2. Main results for model 1.

	Rio de Janeiro 2016	Tokyo 2020+1	Total
n	390	391	781
$\beta_{1,0}$	0.678 (0.291)	0.976 (0.315)	0.817 (0.214)
$\beta_{1,1}$	-0.211 (0.032)	-0.233 (0.034)	-0.221 (0.023)
Null deviance	320.91	321.22	642.14
Residual deviance	211.60	201.76	413.88
AIC	215.60	205.76	417.88

AIC = Akaike Information Criterion.

Table 3. Main results for model 2.

	Rio de Janeiro 2016	Tokyo 2020+1	Total
n	390	391	781
$\beta_{2,0}$	-24.878 (3.067)	-23.203 (2.787)	-23.705 (2.039)
$\beta_{2,1}$	0.012 (0.002)	0.011 (0.001)	0.012 (0.001)
Null deviance	320.91	321.22	642.14
Residual deviance	185.00	179.43	366.40
AIC	189.00	183.43	370.40

AIC = Akaike Information Criterion.

Table 4. Main results for model 3.

	Rio de Janeiro 2016	Tokyo 2020+1	Total
n	390	391	781
$\beta_{3,0}$	-19.770 (4.136)	-16.872 (3.783)	-17.760 (2.749)
$\beta_{3,1}$	-0.061 (0.037)	-0.088 (0.041)	-0.078 (0.278)
$\beta_{3,2}$	0.010 (0.002)	0.008 (0.002)	0.009 (0.001)
Null deviance	320.91	321.22	642.14
Residual deviance	181.85	174.48	357.50
AIC	187.85	180.48	363.50

AIC = Akaike Information Criterion.

Comparing all adjusted and maximal models by LRT, each test provided a p-value of ≈ 1, so we cannot reject the hypothesis that our models are as good as the maximal one. Thus, to choose the best model, we compared them pairwise. Considering AIC, model 2 (AIC = 189.00) is better than model 1 (AIC = 215.60) for Rio, model 2 (AIC = 183.43) is better than model 1 (AIC = 205.76) for Tokyo, and model 2 (AIC = 370.40) is better than model 1 (AIC = 417.88) for both Olympic Games.

Assuming that models 1 and 3, and the models 2 and 3 are nested, LRT was used to compare these models. Models 1 and 3 differed significantly for Rio de Janeiro, Tokyo, and both Olympic Games grouped ($p < 0.001$ for all comparisons). When models 2 and 3 were compared, they did not differ for Rio de Janeiro ($p = 0.076$), but were different for Tokyo ($p = 0.026$), and when both editions were grouped ($p = 0.003$).

As model 3 was better than the others, its quality was assessed by comparing the predicted and real results through a confusion matrix. The cut value was calculated to optimize sensibility and specificity (Table 5).

Table 5. Cut value, accuracy (95% confidence interval), sensitivity (95% confidence interval), specificity (95% confidence interval) and Area Under the Curve (AUC) for model 3.

	Rio de Janeiro 2016	Tokyo 2020+1	Total
Cut value	0.187	0.205	0.210
Accuracy	0.867 (0.829, 0.900)	0.880 (0.843, 0.910)	0.859 (0.833, 0.883)
Sensitivity	0.840 (0.800, 0.602)	0.857 (0.819, 0.888)	0.857 (0.812, 0.863)
Specificity	0.871 (0.834, 0.901)	0.884 (0.848, 0.912)	0.860 (0.853, 0.899)
AUC	0.917	0.927	0.921

Since we have the estimated coefficients, it is possible to calculate de odds ratio OR = exp(β). Finally, the odds ratios for model 3 WLR and Elo coefficients were estimated (Table 6):

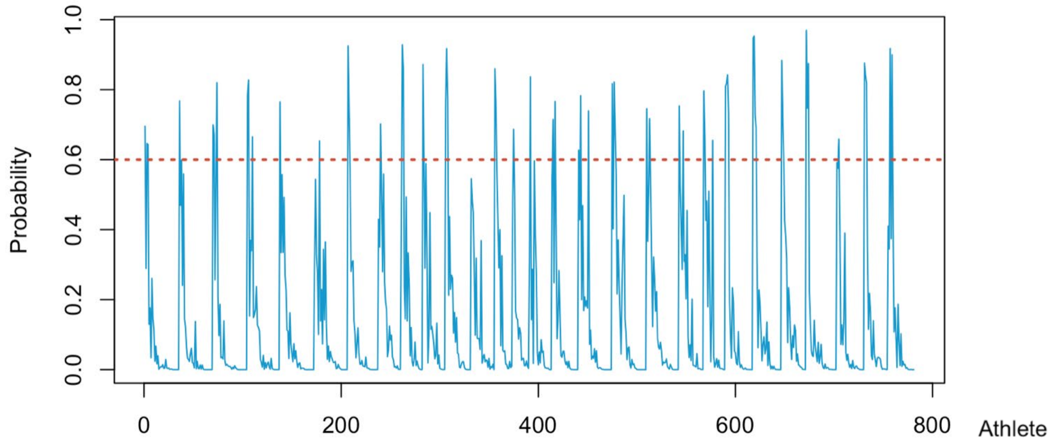
Table 6. Odds ratio (95% confidence interval) for model 3.

	Rio de Janeiro 2016	Tokyo 2020+1	Total
Odds ratio - World Ranking List	0.941 (0.871, 1.006)	0.916 (0.841, 0.991)	0.925 (0.874, 0.975)
Odds ratio - Elo	1.010 (1.006, 1.014)	1.008 (1.005, 1.012)	1.009 (1.006, 1012)

Thus, for each rank position an athlete improves in the IJF WRL, it is expected that there will be an increase in the probability to win an Olympic medal by approximately 7.50%. For instance, for each 10 Elo rating score improvement, it is expected that an athlete will increase the probability to win an Olympic medal in approximately 9.26%.

One limitation in this first approach is that each category has its own estimated probability of success. It led to estimate more than four possible medal winners in a given weight category, while it estimated less than four winners in some other categories. For example, if the cut probability of success is set in 60%, the <70 kg weight category during the Rio Olympics, the shortest peak of Figure 1 below, would not have any predicted medal winner.

Figure 1. Predicted probability of success by athlete.



Therefore, approach 2, performed using model 3 for all weight categories for Rio, Tokyo, and both Olympic Games, was used to solve this question.

3.2. Approach 2

Tables 7, 8 and 9 present the main results for approach 2, model 3.

Table 7. Cut value, accuracy (95% confidence interval), sensitivity (95% confidence interval), specificity (95% confidence interval) and Area Under the Curve (AUC) for model 3 in Rio.

Weight Category	n	Cut value	Accuracy	Sensitivity	Specificity	AUC
Females (kg)						
≤ 48	23	0.463	0.826 (0.629, 0.93)	0.5 (0.311, 0.689)	0.895 (0.709, 0.967)	0.941
≤ 52	22	1	1 (0.851, 1)	1 (0.851, 1)	1 (0.851, 1)	0.951
≤ 57	23	0.459	0.913 (0.732, 0.976)	0.75 (0.547, 0.882)	0.947 (0.777, 0.989)	0.948
≤ 63	26	0.596	0.923 (0.758, 0.979)	0.75 (0.559, 0.876)	0.955 (0.802, 0.991)	0.95
≤ 70	24	0.348	0.75 (0.551, 0.88)	0.25 (0.12, 0.449)	0.85 (0.661, 0.943)	0.943
≤ 78	18	0.637	0.889 (0.672, 0.969)	0.75 (0.519, 0.893)	0.929 (0.722, 0.985)	0.947
> 78	17	1	1 (0.816, 1)	1 (0.816, 1)	1 (0.816, 1)	0.952
Males (kg)						
≤ 60	35	0.467	0.943 (0.814, 0.984)	0.75 (0.587, 0.864)	0.968 (0.85, 0.994)	0.952
≤ 66	34	0.365	0.882 (0.733, 0.953)	0.5 (0.341, 0.659)	0.933 (0.798, 0.98)	0.91
≤ 73	35	0.517	0.943 (0.814, 0.984)	0.75 (0.587, 0.864)	0.968 (0.85, 0.994)	0.912
≤ 81	33	0.306	0.879 (0.727, 0.952)	0.5 (0.339, 0.661)	0.931 (0.793, 0.979)	0.925
≤ 90	35	0.299	0.886 (0.741, 0.955)	0.5 (0.343, 0.657)	0.935 (0.803, 0.981)	0.913
≤ 100	34	1	1 (0.898, 1)	1 (0.898, 1)	1 (0.898, 1)	0.94
> 100	31	0.39	0.871 (0.712, 0.949)	0.5 (0.334, 0.666)	0.926 (0.78, 0.978)	0.942

Table 8. Cut value, accuracy (95% confidence interval), sensitivity (95% confidence interval), specificity (95% confidence interval) and Area Under the Curve (AUC) for model 3 in Tokyo.

Weight Category	n	Cut value	Accuracy	Sensitivity	Specificity	AUC
Females (kg)						
≤ 48	28	1	1 (0.879, 1)	1 (0.879, 1)	1 (0.879, 1)	0.934
≤ 52	29	0.647	0.931 (0.78, 0.981)	0.75 (0.57, 0.872)	1 (0.883, 1)	0.939
≤ 57	25	1	0.84 (0.653, 0.936)	1 (0.867, 1)	1 (0.867, 1)	0.935
≤ 63	31	0.271	0.935 (0.792, 0.982)	0.75 (0.576, 0.869)	1 (0.89, 1)	0.928
≤ 70	28	0.517	0.857 (0.685, 0.943)	0.5 (0.326, 0.674)	1 (0.879, 1)	0.929
≤ 78	24	0.497	0.917 (0.742, 0.977)	0.75 (0.551, 0.88)	1 (0.862, 1)	0.934
> 78	27	1	1 (0.875, 1)	1 (0.875, 1)	1 (0.875, 1)	0.942
Males (kg)						
≤ 60	23	0.36	0.739 (0.535, 0.874)	0.25 (0.118, 0.453)	1 (0.857, 1)	0.789
≤ 66	27	0.444	0.926 (0.766, 0.979)	0.75 (0.563, 0.875)	1 (0.875, 1)	0.866
≤ 73	34	0.522	0.941 (0.809, 0.984)	0.75 (0.584, 0.865)	1 (0.898, 1)	0.931
≤ 81	35	0.33	0.886 (0.741, 0.955)	0.5 (0.343, 0.657)	1 (0.901, 1)	0.906
≤ 90	33	0.408	0.939 (0.803, 0.983)	0.75 (0.582, 0.866)	1 (0.896, 1)	0.903
≤ 100	25	0.436	0.84 (0.653, 0.936)	0.5 (0.318, 0.682)	1 (0.867, 1)	0.907
> 100	22	0.529	0.818 (0.615, 0.927)	0.5 (0.307, 0.693)	1 (0.851, 1)	0.915

Table 9. Accuracy, sensitivity and specificity (95% confidence interval) for model 3.

	Rio de Janeiro 2016	Tokyo 2020+1	Total
Accuracy	0.908 (0.875, 0.933)	0.913 (0.881, 0.937)	0.910 (0.888, 0.928)
Sensitivity	0.679 (0.631, 0.723)	0.696 (0.649, 0.740)	0.688 (0.654, 0.719)
Specificity	0.946 (0.919, 0.965)	0.949 (0.923, 0.967)	0.948 (0.930, 0.961)

4. Discussion

The main goal of the present study was to determine whether the IJF WRL, a ranking created using the Elo system, or another using the combination of both, would better predict Olympic results in judo. The main hypothesis of the present study was that a combination of the WRL and Elo system would predict Olympic performance of judo athletes, which was confirmed. Additionally, for each rank position an athlete improved in the IJF WRL, he/she increased his/her probability to win an Olympic medal by approximately 7.50%, while for each 10 Elo rating score improvement, the athlete increased his/her probability to win an Olympic medal by approximately 9.26%. When both systems were used together, the accuracy of the model was approximately 91%, with a sensitivity of nearly 68-69%, and a specificity close to 95%, for Rio de Janeiro and Tokyo editions isolated or grouped. Such information can serve as a valuable tool for national federations staff in selecting the most suitable athletes to participate in the Olympic judo competition, if both the WRL and an Elo rating system are used together.

Since the introduction of the WRL, several authors have investigated its power to predict judo performance, especially at Olympic level (Brunel, 2022; Daniel & Daniel, 2013; Franchini & Julio, 2015; Guilheiro & Franchini, 2017; Santos et al., 2023), using different approaches.

Daniel and Daniel (2013) were the first to use the WRL, WRL-derived variables (e.g., WRL top 8 athletes), and previous competitive performance (e.g., Beijing Olympic Games and 2011 World Championship performances) to verify their predictive power regarding London Olympic Games results. They considered only the medal winners in their analysis, and reported that 81% of the medal winners were positioned among the top 8 in the WRL, and 70% of the medal winners were top 8 in the previous World Championship, when the two half-lightweight categories (<52 kg for females, and <66 kg for males) were not included. Therefore, these authors conducted only descriptive analysis and did not consider all weight categories. Franchini and Julio (2015) analyzed the same Olympic Games edition (i.e., London 2012) performance of judo athletes and indicated that variables derived from the WRL (e.g., points valid in the two years preceding the Olympic Games, number of competitions disputed in the year of the Olympic Games, percentage of matches won in the year of the Olympic Games) were able to predict 50-51% of final position in that competition. Thus, these authors considered all competitors, but used only multiple linear regression in their analysis. Similarly, Santos et al. (2023) applied multiple linear regression analysis and reported that the percentage of matches won during the classification period and competition in the year prior to the Olympic Games predicted 37% of female judo athletes performance in Tokyo Olympics. For male judo athletes, they observed that the percentage of matches won during the classification period and competition in the six months before the Olympic Games predicted 36% of performance. Therefore, regarding Olympic Games performance prediction only descriptive or linear approaches were used, which may have limited the detection of the predictive value of previous performance.

The second most important competition in judo is the World Championship, which is disputed for athletes from the cadet category and older ages. Breviglieri et al. (2018) investigated the performance in the World Championships for cadets, juniors, and seniors disputed in 2017. Like in other studies, these authors also used WRL-derived variables to predict performance in the World Championships. They observed that WRL and short-term performance (i.e., performance in competitions close to the World Championship) predicted between 5% and 27% of the final position in the World Championships for athletes in these age groups. Specifically, they reported that: (1) for the senior age group, those among the top-ranked athletes in the draw, but who performed a lower number of competitions, were factors associated with better performance; for juniors, being among the top-ranked athletes in the draw (for males), being the best-ranked athlete, and presenting a higher winning percentage in the year of the competition (for females), better predicted performance; for cadets, a higher number of wins up to the World Championship, but a lower number of total matches up to this event, were the main factors associated with performance. These authors also used multiple linear regression, which may have limited their findings.

The ranking system has also been used for national-level competitions, and some authors (Courel-Ibañez et al., 2018) investigated the prediction of the ranking position on competitive performance. They reported that high-ranked athletes presented a higher probability of winning a match and advancing to the next phase, and this advantage was more significant at the initial phases for female judo athletes but mitigated as the championship progressed. Another important finding of this study was that the quarter-finals were a critical phase for senior male judo athletes, in which better-ranked athletes were more likely to defeat lower-ranked ones. Junior male judo athletes did not present any advantage of being better ranked in terms of performance in a national championship. For senior females, the advantage was more pronounced in the elimination and semi-final phases, while for junior females a higher probability of winning was present mainly in the elimination and quarter-finals. This approach is relevant, as it considers different phases of the competition and the probability of winning based on ranking position.

One important feature of the WRL is the positioning of athletes as seeded or not seeded in competition (Brunel, 2022; Guilheiro & Franchini, 2017). Using a Bayesian approach, Guilheiro and Franchini (2017) investigated whether being seeded affected the probability of winning a medal during the London and Rio de Janeiro Olympic Games editions. They reported that the probability of winning a medal was 41.1% and 42.9% for male seeded athletes, and of 35.7% and 44.6% for female athletes for the 2012 and 2016 editions, respectively. For instance, Brunel (2022) applied Monte-Carlo simulations to estimate the probability of winning a competition, reaching the final, or winning a medal in a standard draw compared to a random one. This author indicated that being seeded had limited advantage to win the competition but would positively affect the likelihood of winning a medal. Conversely, athletes with a high quality but misclassified in the WRL (e.g., an athlete with a high percentage of winning but with a small number of competitions) would be at a disadvantage merely due to the seeding process.

Taken together, these studies demonstrate that when employing multiple linear regression approaches using the WRL or WRL-derived variables, the prediction of Olympic performance is low. However, as the WRL determines the seeding of athletes, more sophisticated analyses indicate that these athletes are more likely to win. Therefore, the main contribution of adding the Elo rating system to predict judo athletes' Olympic performance is likely related to incorporating opponent quality into account (Minton, 2017), which is a key element in an event where only the best athletes can compete. Even though the WRL may capture the consistency and ability that a given athlete has during the two previous years of the Olympic Games, it does not consider that some athletes may be wise in selecting competition and inflating his/her ranking position due to the lack of the best opponents. Thus, using a model that considers both the WRL and the Elo rating system was able to better predict Olympic performance. Consequently, judo managers and stakeholders must consider determining the Elo rating of their judo athletes and base their decision regarding athletes selection on a combination of the WRL and the Elo rating system when two athletes from the same nation are in the classification zone for the Olympic Games.

The main limitation of the present study is related to the fact that the sample size is limited to number of athletes qualified for the Olympic Games and to the schedule of competitions they followed to achieve the points needed for this qualification. Additionally, the Elo Rating System is a model in itself, so it has its own limitations. In this study, we used a constant k-factor for each weight category, but it could vary according to the quality of competition, proximity to the Olympic Games, and other factors. Another limitation of the Elo System arises from the reality of the IJF World Tour. Athletes do not compete in every event, leading to a disconnection between the time period and the number of events in which each athlete participates. The Glicko System (Glickman, 2016) may address such issues, as it penalizes the ratings of athletes who compete less, estimating its uncertainty.

5. Conclusions

When employing models that utilize the WRL, the Elo rating system, and a combination of both to predict Olympic-level judo performance, the combined model yielded better estimates. Additionally, for each rank position an athlete improved in the IJF WRL, there was approximately a 7.5% increased probability of winning an Olympic medal, while for each 10 Elo rating score improvement, a roughly 9.2% increased probability to win a medal was observed. When both systems were combined, the accuracy of the model was approximately 91%, with a sensitivity of 68-69%, and a specificity close to 95%, for both isolated Rio de Janeiro and Tokyo editions, or when grouped together.

References

Belmont Report. (1979). https://videocast.nih.gov/pdf/ohrp_belmont_report.pdf

Breviglieri, P. V., Possa, M. E. S., Campos, V. M., Humberstone, C., & Franchini, E. (2018). Judo world ranking lists and performance during cadet, junior and senior World Championships. *Ido Movement for Culture. Journal of Martial Arts Anthropology*, *18*(2), 48–53. http://doi.org/10.14589/ido.18.2.7

Brunel, V. (2022). Seed advantage in sport competitions: the case of professional judo. *Revista de Artes Marciales Asiáticas*, *17*(2), 108-118. http://doi10.18002/rama.v17i2.7047

Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). *Journal of Royal Statistical Society Series B*, *20*(2), 215–242. https://www.jstor.org/stable/2983890

Courel-Ibáñez J., Escobar-Molina R., & Franchini E. (2018). Does the ranking position predict the final combat outcome in senior and junior judo athletes? *Revista de Artes Marciales Asiáticas*, *13*(2), 131-138. http://dx.doi.org/10.18002/rama.v13i2.5471

Daniel, L.F., & Daniel, R. (2013). Study regarding the prediction of medal winning in Olympic Games judo competitions. *Journal of Physical Education and Sport*, *13*(3), 386-390. https://doi.org/10.7752/jpes.2013.03062

Elo, A. E. (1978). *The rating of chess players, past and present*. Arco.

Franchini, E., & Julio, U.F. (2015). The judo world ranking and the performances in the 2012 London Olympics. *Asian Journal of Sport Medicine*, *6*(3), e24045. https://dx.doi.org/10.5812%2Fasjsm.24045

Franchini, E., Takito, M.Y., da Silva, R.M., Shiroma, S.A., Wicks, L., & Julio, U.F. (2017). Optimal interval for success in judo world-ranking competitions. *International Journal of Sports Physiology and Performance*, *12*(5), 707-710. https://doi.org/10.1123/ijspp.2016-0375

Glickerman, M. (2016). *The Glicko system*. Retrieved from http://www.glicko.net/glicko/glicko.pdf

Guilheiro, L. M. (2020). *Modelagem matemática do ranking mundial de judô e do desempenho relativo entre atletas de elite*. [Monograph, Unisul, Palhoça, SC]. Supervisor: Christian Wagner.

Guilheiro, L. M., & Franchini, E. (2017). Be seeded or not be seeded? A study with Olympic judo athletes. *Journal of Exercise Rehabilitation*, *13*(2), 148-152. https://doi.org/10.12965/jer.1734904.452

International Judo Federation (2024). Sport and Organisation Rules. www.ijf.org. 15 April 2024.

International Judo Federation (2024). Qualification System - Games of the XXXIII Olympiad - Paris 2024. https://78884ca60822a34fb0e6-082b8fd5551e97bc65e327988b444396.ssl.cf3.rackcdn.com/up/2023/12/2023-11-28_-_Olympic_Games_Par-1701595981.pdf 15 April 2024.

Langville, A. N., & Meyer, C. D. (2012). *Who's #1? The science of rating and ranking*. Princeton University Press.

Minton, R.B. (2017). *Sports math - an introductory course in the mathematics of sports science and sports analytics*. CRC Press.

Neumann, C., & Kulik, L. (2020). *Animal dominance hierarchies by Elo Rating* [R]. Retrieved from https://github.com/gobbios/EloRating

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers and Education*, 98 ,169-179. https://doi.org/10.1016/j.compedu.2016.03.017

Santos, D. F. C., Kons, R. L., Lopes-Silva, J. P., Agostinho, M. F., Detanico, D., Takito, M. Y., & Franchini, E. (2023). Participation in the International Judo Federation World Tour competitions and performance in Tokyo Olympic Games. *Frontiers in Sports and Active Living*, *27*(5), 1216002. https://doi.org/10.3389/fspor.2023.1216002

Wicks, L (2011). An Experimental Relative Skill Based Ranking System for Elite Level Judo. *Abstracts of the 7th IAJR International Judo Research Symposium*. Retrieved from: http://judoresearch.org/wp-content/uploads/2012/03/Lance-Wicks.pdf access on 15th April 2024.

~

Author's biographical data

Leandro Marques Guilheiro (Brazil) is a Master's student in Mathematics at the Institute for Pure and Applied Mathematics (MPA), where he investigates sports analytics. Currently, he serves as the head coach of the Esporte Clube Pinheiros judo team. Under his guidance, his athletes have won medals at the Olympic Games and World Championships across cadet, junior, and senior age categories. He holds a 6th dan in judo. As an athlete, he was a Junior World Champion, a double Olympic bronze medalist (Athens 2004 and Beijing 2008), and a double Senior World Championship medalist (silver in Tokyo 2010 and bronze in Paris 2011). E-mail: lguilheiro@hotmail.com

Emerson Franchini (Brazil) is a full professor in the Department of Sport at the School of Physical Education and Sport, University of São Paulo, where he coordinates the Martial Arts and Combat Sports Research Group. He has been conducting research on combat sports since 1995, with a particular focus on judo. He holds a 2nd dan in judo and has worked as a consultant for international-level judo athletes, including the first author, who have won medals at the Olympic Games and World Championships. E-mail: efranchini@usp.br