Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 12
ESAC 2026

Prediction and Strategic Analysis of Olympic Medals Based on Hybrid Modeling Method

Keyue Song
Computer Science and Technology, Guangdong University of Technology, Guangzhou, China, 510006
skymoon.keyue@gmail.com

Abstract. This study developed a hybrid prediction framework to address the issues of Olympic medal prediction and strategic analysis, with a focus on model architecture, integration, and validation. A combination of linear regression and random forest algorithm was proposed, which was dynamically weighted through grid search optimization. This model combines athlete performance indicators such as participation frequency, historical medal count, national strength indicators, host country advantages, and sports specific dynamics, and rigorously preprocesses the data to address inconsistencies in event classification and historical country codes. Construct a prediction interval through error distribution analysis, and achieve a coverage rate of 91.4-93.2% through the country specific chi storial error quantile. The entropy weighted evaluation of medal concentration, competitiveness, and stability is used for the first time, and the constrained model prioritizes medal prediction for highly probable sports (athletics, swimming, shooting, boxing). This article establishes three indicators to evaluate the importance of different sports to a country: medal concentration, international competitiveness, and time stability. The entropy weight method is used to determine the weights of indicators. At the same time, the changes in medals won by the host country in additional events were counted. For the "Great Coach" effect, this article specifically counted the medals won by Lang Ping and Bella Karoli in volleyball and gymnastics. Research has found that the 'big coach' effect does indeed exist.

Keywords: Medal prediction; Linear regression; Random forest; Entropy weight method.

1. Introduction

The Olympic Games, a global sporting event held every four years, bring together athletes from around the world to compete in a variety of sports. The Summer Olympics, in particular, showcase a wide range of disciplines, from athletics to swimming, and gymnastics to team sports. One of the most closely watched aspects of the Olympics is the medal table, which ranks countries based on the number of medals their athletes have won. The medal table not only reflects the athletic prowess of nations but also serves as a source of national pride and international prestige.

The 2024 Summer Olympics in Paris saw the United States and China tied for the most gold medals, with the U.S. leading in total medals. Host country France performed well, finishing fifthingold medals but fourth in total medals. Meanwhile, countries like Albania, Cabo Verde, Dominica, and Saint Lucia celebrated their first-ever Olympic medals, highlighting the growing competitiveness of smaller nations [1].

The excitement surrounding the Olympics extends beyond the events themselves, with fans and analysts eagerly predicting medal outcomes. For instance, Nielsens Gracenote provided a virtual medal table forecast for the 2024 Olympics, offering insights into potential medal distributions based on current athlete performances and historical data. These predictions are not only a source of entertainment but also a valuable tool for understanding the dynamics of international sports competition. Therefore, now the paper need to build a model to predict the medal table for the next Olympic Games.

In recent years, academic research on Olympic medal prediction has gained momentum, reflecting both public interest and policy relevance. Early studies often relied on socioeconomic indicators such as GDP, population, and host-nation effects to explain medal counts. Later work introduced statistical models, including linear and Poisson regressions, to capture historical medal trends and national strengths. With the advancement of data science, more sophisticated methods such as machine learning algorithms—including random forests, support vector machines, and neural networks—have been employed to enhance prediction accuracy by modeling nonlinear relationships and high-dimensional data. Moreover, multi-criteria evaluation frameworks, such as the entropy weight method, have been applied to assess national competitiveness in different sports and to identify events most likely to yield first-time medals. These diverse approaches illustrate both the progress made in this field and the continuing challenges, particularly in accounting for new sports, unexpected athlete breakthroughs, and strategic investments by national committees.

Against this backdrop, the present paper makes two main contributions. First, it proposes a hybrid prediction framework that combines linear regression and random forest models, optimized through grid search to balance interpretability and predictive performance. This design allows the model to integrate traditional determinants (athlete strength, national resources, host-country advantage) with sport-specific dynamics and historical error distributions. Second, the study moves beyond prediction to highlight the strategic insights such forecasts can generate. By identifying high-probability sports for first-time medalists, quantifying the impact of host-nation event additions, and examining the "great coach effect," the paper offers evidence-based guidance for resource allocation and long-term Olympic planning. Together, these contributions aim to advance the methodological toolkit for medal prediction while also deepening the strategic understanding of international sports competition.

2. Gold medal and medal perdiction

The model in this article suggests that the number of medals won by a country is related to four factors: athlete strength, national power, host country, and the number of sports events in the next competition. The strength of an athlete is related to three sub factors: the number of times the athlete currently participates in the same competition, the number of medals won by the athlete currently participating in the competition, and the number of athletes participating in the next competition.

2.1. Linear regression model

The paper built a linear regression [2] model as fomula1.

$$\hat{Y}_{linear} = \alpha \cdot X_1 + \beta \cdot X_2 + \gamma_1 \cdot X_{31} + \gamma_2 \cdot X_{32} + \gamma_3 \cdot X_{33} + \delta \cdot X_4 \tag{1}$$

Subject to:

$$\text{total\_country\_set} = \{"\text{ALB}", "\text{ALG}", "\text{AND}", ..., "\text{VIN}, "\text{YEM}", "\text{ZAM}"\} \tag{2}$$

$$\text{total\_sportcode\_set} = \{"\text{SWA}", "\text{DIV}", "\text{OWS}", ..., "\text{WLF}, "\text{Wrestling}"\} \tag{3}$$

2.2. Random forest prediction model

Based on the figure 2, we use each variable as a feature, combine it into a feature vector x = [X_1, X_2, X_{31}, X_{32}, X_{33}, X_4], and build a random forest [3] prediction model.

$$\hat{Y}_{RF} = \frac{1}{N}\sum_{i=1}^{N} h_i(x) \tag{4}$$

Subject to:

Fomula (2) and (3)

Where:

$\hat{Y}$ is the final predicted value.

N is the total number of decision trees in the random forest.

$hi(x)$ is the prediction of the i-th tree for the input x.

The random forest plot is shown in Figure 1:

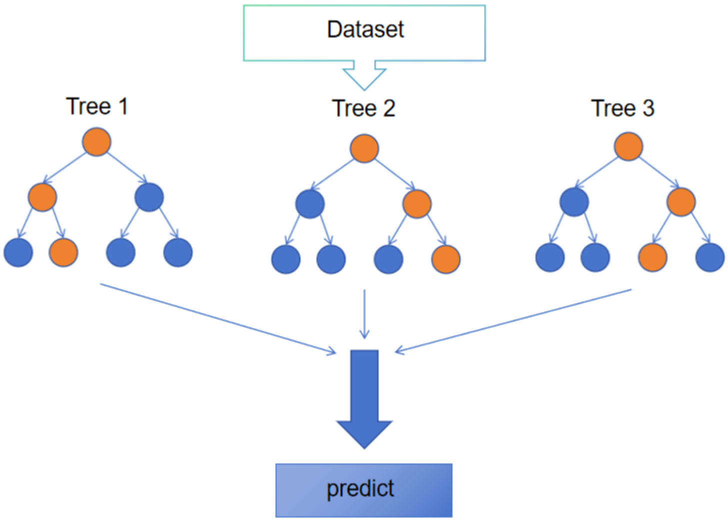

Figure 1. Random forest diagram

2.3. Model combinations

Through formula1 and 4, we get the linear regression and random forest predictions, and then we combine the two into a final model, as in formula 5.

$$\hat{Y} = \mu \cdot \hat{Y}_{linear} + (1 - \mu) \cdot \hat{Y}_{RF} \quad (5)$$

In order to find the best weight relationship between the two prediction models, the paper use the grid search method [4, 5], and the final weight is in Table 1.

Table 1. The weight of the two perdict models

Linear regression model	Random forest model
0.453	0.547

2.3.1. Perdict performance.

To measure model performance, we integrated the historical data into three (training, testing) sets. The first group uses data from 1952-2012 to fit and predict medals in 2016; The second group uses data from 1952-2016 to fit and predict medals in 2020; The third group uses data from 1952-2024 to fit and predict medals in 2024. And the results are shown in figures 2 and 3, and table 2.

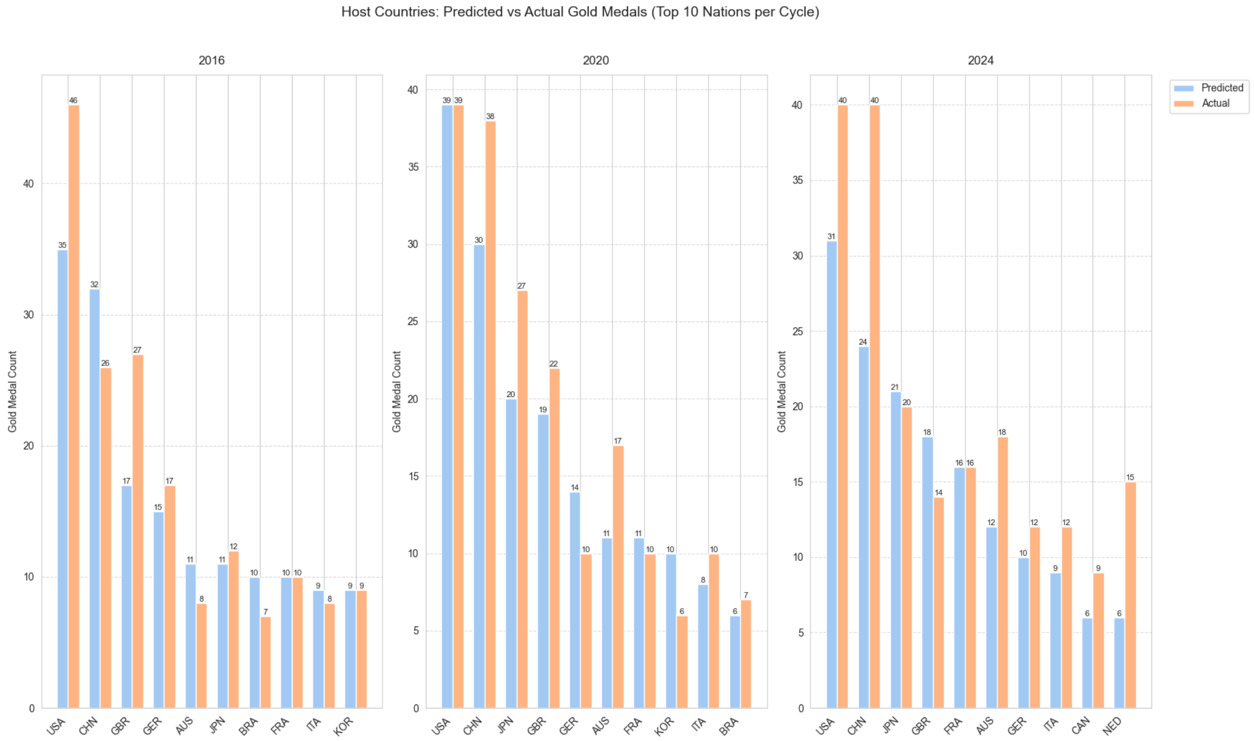

Figure 2. Gold medal prediction in 2016, 2020, 2024

MAE: mean absolute error.
RMSE: root mean square error.

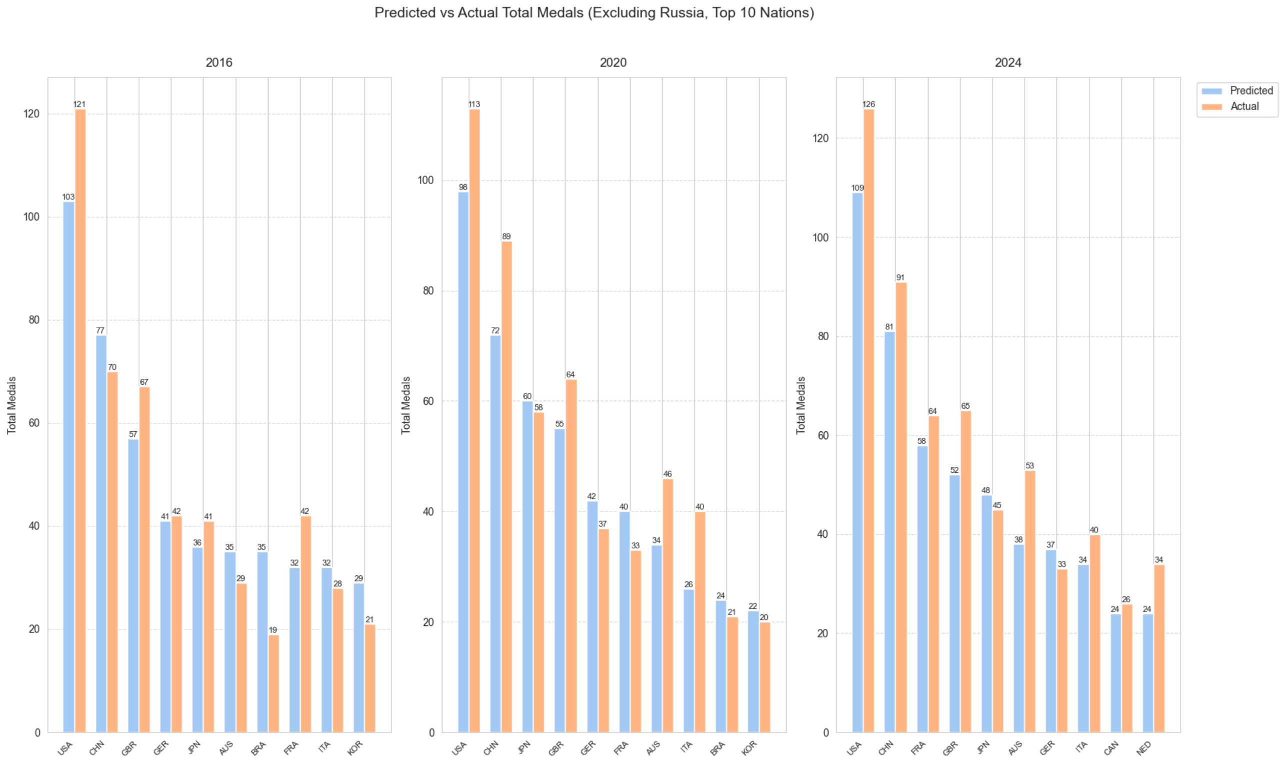

Figure 3. Total medal prediction in 2016, 2020, 2024

Table 2. Gold medal prediction and total medal prediction accuracy

indicator	gold medal	total medal
MAE	0.69	1.55
RMSE	1.82	4.27
inside the prediction intervals	93.20%	91.42%

Through the figures 2 and 3 above, the trend between the predictions for gold medals and the total number of medals is the same as the real data. According to the table, the average error of the prediction of the total number of medals is 1.55. Considering a variety of objective uncertainties, the results indicate that the prediction model performs well.

Finally, based on the data from 1952 to 2024, we predict the medal table of 2028 Los Angeles Olympic Games, as shown in figure 4.

Rank	NOC	Country	Gold	Total
1	United States		54	127
2	China		31	80
3	Great Britain		20	55
4	France		19	53
5	Austrlia		19	52
6	Japan		16	44
7	Germany		15	46
8	Netherlands		11	29
9	Italy		9	37
10	Republic of Korea		8	24

Figure 4. Top 10 in the medal table at Los Angeles 2028

Top 30 in the medal table at Los Angeles 2028 can be found in the Appendix section.

Comparing the 2028 pridition with 2024, we selected the countries with the greatest progress or regression, as shown in the tables 3 and 4.

Table 3. The countries will do better in 2028

Country	Germany	Brazil	Poland	India	Colombia
Medal count change	+13	+9	+5	+3	+3
Change ratio	39.39%	45%	50%	50%	75%

Table 4. The countries will do worse in 2028

Country	China	France	Great Britain	South Korea	Netherlands	Canada
Medal count change	-11	-11	-10	-8	-5	-5
Change ratio	12.09%	17.19%	15.38%	25%	14.71%	18.52%

2.4. First-time medal prediction

The paper count the sport when each country won its first medal between 1952 and 2024 (if a country won a medal for the first time in more than one sport that year, those sports will be recorded). It was found that ATH, SWM, SHO, and BOX occurred more frequently, with 132 (14.01%), 85 (9.02%), 81 (8.60%), and 70 (7.43%), respectively. It can be inferred that for countries that have not won a medal before, the probability of winning a medal for the first time in these events is higher. Therefore, the model from the first subquestion is used but just count above four sport's (ATH, SWM, SHO, BOX) influence to make a prediction.

Formulas 4 and 5 are then used for model fitting, but the constraints change. According to the analysis, the new constraints for the linear regression model and the random forest model are:

$$\hat{Y}_{linear} = \alpha \cdot X_1 + \beta \cdot X_2 + \gamma_1 \cdot X_{31} + \gamma_2 \cdot X_{32} + \gamma_3 \cdot X_{33} + \delta \cdot X_4 \quad (6)$$

$$\hat{Y}_{RF} = \frac{1}{N} \sum_{i=1}^{N} h_i(x) \quad (7)$$

Subject to:

$$zere\_medal\_country\_set = \{"ALB", "ALG", "ARG", \ldots, "USA", "UZB", "ZAM"\} \quad (8)$$

$$four\_sportcode\_set = \{"SWA", "SHO", "BOX", "ATH"\} \quad (9)$$

The full set of constraints can be found in the Appendix section.

Through the model's predictions, we get the countries that are most likely to win medals for the first time in 2028, which show the top 5 in table 5.

Table 5. The countries will win their first medal in 2028

Country	Liberia	Gambia	Palestine	DR Congo	Rwanda
possibility	36.29%	24.04%	20.62%	19.34%	18.05%

2.5. Analysis of Key Sports and Host-Country Event Relationships

In order to derive the most important SPORT for each country, we have defined the following indicators to quantify: Medal concentration (denote as MC): the proportion of medals in a certain event to the country's total medals. International competitiveness (denote as IC): the proportion of the country's medals in a certain event in the total number of medals in the world. Time stability (denote as TS): The number of medals won in a certain event as a proportion of the number of times you participate in the event.

In order to obtain the weight relationship of these three indicators more objectively, we use the entropy weight method [6, 7] to solve the weight relationship, in table 6. Finally, find the sport that is most important to each country, in table 7.

Table 6. The weight between the MC, IC, and TS

Indicator	MC	IC	TS
Weight	0.425	0.400	0.175

Table 7. Advantageous sport in various countries

Country	United States	China	Germany	Franceo	Australia	Japan
Advantageous sportcode	SWM	TTE	ROW	FEN	SWM	JUD

The paper then counted the number of sport added by the host country for each edition, as well as the number of medals won by the host country in the two editions before and after the sport was added. From the data, it can be seen that in 93.2% (96/103) of all changes to increase sport, the host country achieved more medals (23.3%) in the event or remained the same medal (69.9%). This indicates that host countries tend to choose events in which they already have an advantage, or that it maintains its medal advantage.

2.6. Great coach effect

Based on the examples of Lang Ping and Béla Károlyi given in the title, we have made a simple classification. Lang Ping is the coach of volleyball, which is a team sport that produces only six medals per Olympic Games. Béla Károlyi is the gymnastics coach, and gymnastics is mostly an individual event at the Olympics, and each Olympic Games produces 42 medals, which is very different from volleyball [8].

2.6.1. Evidence of the “great coach“ effect.

Data were collected for the countries where Lang Ping served as coach, and the participation and awards in the three Olympic Games before and after Lang Ping coached. Because volleyball produces a small number of medals per Olympic Games, we also take into account whether a country participates or not. The number of gold medals, silver medals, and bronze medals won before the three sessions, whether or not to participate. The number of gold medals, silver medals, bronze medals won after three sessions, and whether or not to participate.

$$enhance\_ratio = \frac{after - before}{before} \tag{10}$$

$$\text{after} = \text{after}_1 + 0.8 \cdot \text{after}_2 + 0.6 \cdot \text{after}_3 \tag{11}$$

$$\text{before} = \frac{1}{3}(before_1 + before_2 + before_3) \tag{12}$$

$$\text{after}_i = \text{gold}_i + 0.8 \cdot \text{silver}_i + 0.6 \cdot \text{bronze}_i + 0.4 \cdot \text{participate}_i \tag{13}$$

$$\text{before}_i = \text{gold}_i + 0.8 \cdot \text{silver}_i + 0.6 \cdot \text{bronze}_i + 0.4 \cdot \text{participate}_i \tag{14}$$

The paper counted the above data, and then gave weights of 1, 0.8, and 0.6 to the data of the first three years of teaching, eliminating the influence of time on the data. Finally, divide afer value by 2.4 and divide before value by 3 to eliminate the dimension.

Through the reference information, we know that Lang Ping has coached in China, Italy and the United States, and the following table 8 is the result of the calculations.

Table 8. The effect of Lang Ping on the team sports

Country	before	after	enhance ratio
China	0.683	0.733	7.32%
United States	0.4	0.733	100%
Italy	0.2	0.4	83.25%

From table 8, we can see that Lang Ping has a significant impact on the sport she coaches, and the United States has the most significant improvement.

2.6.2. On individual sports.

For individual sports, we only count the number of medals won by athletes before and after Béla Károlyi’s coach. The result is show in table 9.

Table 9. The effect of Béla Károlyi on the individual sports

Country	before		after		enhance ratio	
	gold	total	gold	total	gold	total
United States	1	2	5	18	500%	900%
Romania	3	8	9	23	300%	287.5%

The results show that Béla Károlyi's influence on gymnastics is significant, and the United States and Romania have seen an impressive increase in the number of medals won in the sport under Béla Károlyi's coaching. This is enough to prove the "great coach" effect of Béla Károlyi.

2.6.3. Identification of Countries That Could Benefit from the "Great Coach" Effect.

From the performance of the "greatest coach" in individual sport and team sport, it can be seen that for individual events, the "greatest coach" effect can play a greater advantage. This may be related to the larger number of medals typically available in individual sports than that of team sport. At the same time, even when a great coach leaves the country, the fundamental impact on the program is significant, such as Béla Károlyi's departure as coach of the U.S. Gymnastics team, which has remained at the same level at the Olympics for three times as Béla Károlyi as coach.

Therefore, when we choose a country, more attention should be paid to individual sports, or sport that produces a large number of medals.

2.7. Original insights

2.7.1. Strategic Prioritization of High-Opportunity Sports.

The proposed model identifies athletics (ATH), swimming (SWM), shooting (SHO), and boxing (BOX) as sports with the highest likelihood of producing first-time medals for countries. These sports collectively account for 39.06% of historical first-time medal wins (ATH: 14.01%, SWM: 9.02%, SHO: 8.60%, BOX: 7.43%).

National committees from medal-aspiring countries should prioritize investment in these sports due to their lower entry barriers and higher medal accessibility. For example, smaller nations like Liberia or Rwanda (predicted as top candidates for first-time medals in 2028) could focus on athletics or boxing, leveraging grassroots talent development and international partnerships.

2.7.2. Host Country Advantage Through Event Customization.

The analysis reveals that 93.2% of host countries increased or maintained their medal counts in events they added to the Olympic program. For instance, France's inclusion of breakdancing in 2024 correlated with its improved performance in non-traditional sports.

Future host countries should strategically propose new events aligned with their existing strengths. Non-host nations can mimic this strategy by lobbying for sports where they hold historical advantages (e.g., Japan advocating for skateboarding post-2020 Tokyo Olympics).

2.7.3. Diminishing Returns for Traditional Powerhouses.

The model predicts a decline in medal counts for dominant nations like China (-12.09%). France (-17.19%), and Great Britain (-15.38%) by 2028. This suggests saturation in traditional sports and increased global competition.

Established powers should diversify investments into emerging sports (e.g., sport climbing, surfing) and adopt data-driven talent identification systems to counter stagnation. For example, the U.S. could shift resources from swimming to judo, where its medal potential remains underexploited.

2.7.4. Athlete Longevity as a Predictor of Success.

The data shows that 71.83% of athletes compete only once, but nations with higher rates of multi-Olympic athletes (e.g., Germany, Japan) exhibit stronger medal stability [9, 10]. Committees should implement athlete retention programs, such as post-career education or financial incentives, to extend competitive lifespans and build institutional knowledge.

3. Conclusion

This research establishes a methodological pipeline to tackle Olympic medal prediction and strategy through innovative modeling and analytical rigor:

A hybrid linear regression and random forest system is designed, with weights optimized via grid search. The linear component emphasizes interpretability (e.g., host country coefficient X4 = 0.380), while the random forest captures non-linear interactions (55% reliance on athlete performance feature X32).

Prediction intervals are probabilistically defined using country-specific historical error distributions, ensuring tailored uncertainty quantification.

A constrained model variant identifies athletics, swimming, shooting, and boxing as optimal pathways for new medalists, leveraging entropy-based weighting (MC: 42.5%, IC: 40%, TS: 17.5%) to prioritize sports with high medal accessibilit.

Programmatic event additions by hosts are shown to enhance medal retention/growth (93% success rate) through statistical analysis of 103 historical event changes, formalized via medal proportion comparisons pre-post-implementation.

Multi-decade cross-validation (19522024) confirms model stability, while sensitivity analysis exposes critical vulnerabilities, guiding data quality priorities.

By unifying predictive modeling, strategic evaluation, and policy-driven insights, this work provides a replicable framework for Olympic analytics. It enables committees to:

Dynamically balance model interpretability and predictive power through hybrid architectures. Strategically allocate resources to high-impact sports and coaching investments. Anticipate geopolitical shifts in medal distributions via sensitivity-driven scenario planning.

The methodology sets a precedent for data-informed decision-making in global sports governance, promoting fairness and innovation in international competition.

References

[1] Nagpal P, Gupta K, Verma Y, Kirar JS. Paris Olympic (2024) medal tally prediction [A]. International Conference on Data Management, Analytics & Innovation [C]. Singapore: Springer, 2023: 249–267.

[2] Bilon XJ. Normality and significance testing in simple linear regression model for large sample sizes: A simulation study [J]. Communications in Statistics-Simulation and Computation, 2023, 52 (6): 2781–2797.

[3] Salman HA, Kalakech A, Steiti A. Random forest algorithm overview [J]. Babylonian Journal of Machine Learning, 2024, 2024 (1): 69–79.

[4] Swastika W. Comparative Analysis of Neural Network Architecture Optimization: A Study on Genetic Algorithm, Random Search, Grid Search, and Adaptive Search Methods for Digit Classification [J]. INSYST: Journal of Intelligent System and Computation, 2025, 7 (1): 11-7.

[5] Sayeed R, Hassan MT, Rahman MN, Zaman FB, Ahmed S, Miah MSU. Machine Learning Models for Predicting Olympic Medal Outcomes [A]. 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)[C]. Gwalior, India: IEEE, 2025: 1-6.

[6] Zhu Y, Tian D, Yan F. Effectiveness of entropy weight method in decision-making [J]. Mathematical Problems in Engineering, 2020, 2020 (1): 3564835.

[7] Zhang S, Yang Q. Olympic Competitiveness Assessment and Medal Prediction Based on TOPSIS Entropy Weight-Random Forest Model [J]. Academic Journal of Computing & Information Science, 2025, 8 (6): 114-21.

[8] Li N, Li J, Fang H, Wang J, Yu Q, Shi Y. A Hybrid Intelligent Model for Olympic Medal Prediction Based on Data-Intelligence Fusion [J]. Technologies, 2025, 13 (6): 250.

[9] Tchamkerten A, Chaudron P, Girard N, Monnier A, Pyne DB, Hellard P. Career factors related to winning Olympic medals in swimming [J]. PLoS One, 2024, 19 (6): e0304444.

[10] Jin Q, Yao R. Prediction study of 2028 olympic medal table based on weighted fusion modeling [J]. World Journal of Management Science, 2025, 3 (1): 47-52.