Olympic Potential Mining: Ensemble Learning & Multi-Criteria Decision-Making

Wei Zhong[1,a], Yan Zhang[2,b], Zichen Zhao[3,c]

[1] Central South University, Changsha, 410083, China;

[2] Shaanxi Normal University, Xi'an, 710119, China;

[3] The University of Melbourne, Melbourne, 3010, Australia;

Wei Zhong, Yan Zhang and Zichen Zhao are the co-first authors

Abstract. With the global focus on Olympic performances rising, predicting medal distributions and evaluating coaching strategies are vital for national sports planning. This paper builds a comprehensive modeling framework using historical Olympic data to solve these problems. A Pearson correlation-based descriptive statistical analysis shows weak feature correlations. Second, the data is trained with models like Random Forest (RF), Gradient Boosting Decision Tree (GBDT), CatBoost, and a Stacking ensemble model. After comparison, the Stacking-based model is chosen for prediction. It outperforms single algorithms in accuracy. A Random Forest - EDA analysis reveals that participating in competitive events like swimming and athletics boosts medal counts. For every 8% increase in event diversity, medals rise by 7.2% on average. Host countries can gain 10 - 15% more medals by choosing advantageous events, like France adding wrestling and handball. Next, the SHAP - EWH - TOPSIS framework quantifies the "Great Coach" effect. SHAP values show that a coach's experience and number of elite athletes trained are key in medal prediction. The Entropy Weight-modified TOPSIS model ranks coaches. Pep Guardiola and Lang Ping are at the top. Finally, sensitivity analysis validates model robustness. Medal predictions are most sensitive to the number of participating events and the coaches' athlete-training ability. The EWH-TOPSIS framework is stable with ±20% weight fluctuations, and coach rankings have 92% consistency. This paper offers practical guidance for National Olympic Committees in resource allocation, coach recruitment, and event participation strategy. The machine - learning and multi - criteria - decision - making model framework is a reliable tool for predicting Olympic success and unlocking sports potential.

Keywords: Olympic medal prediction; Stacking ensemble learning; SHAP - EWH - TOPSIS.

1. Introduction

The Olympic Games, as the most influential sporting event in the world, bring together athletes from across the globe, showcasing the sporting prowess and developmental levels of various nations. The medal table, serving as a crucial indicator of a country's athletic achievements, attracts significant attention. With the increasing professionalization of sports competitions, the performance of countries in the Olympics is not only influenced by the athletes' abilities but is also constrained by numerous external factors, such as the setup of events, the impact of coaching teams, and national sports development policies. Therefore, establishing an accurate medal prediction model that integrates historical data, athlete performance, event characteristics, and other multifaceted factors can provide a scientific basis for countries to formulate their Olympic preparation strategies.

A broad spectrum of academic attention has been given to studies on Olympic medals in recent years. The primary focus in most studies is on building predictive models with historical information that can predict the winners of Olympic medal competitions in the future. Many of the models are constructed using statistical tools and machine learning tools like linear regression (the most popular example being Haysopter), random forest, Great Fun Run (FM), and deep learning networks, which help in increasing prediction accuracy and confidence [1]. Apart from the specific athletic performance of athletes, evidence suggests that other macroeconomic factors, such as the design of stadium facilities and socioeconomic contexts, also contribute to medal expectations, including individual athlete performance [2]. Models commonly consider estimating uncertainties and quantifying prediction accuracy by using confidence intervals, which are crucial in examining how different factors affect outcomes [3]. Supported by new research indicating that coaches' transnational mobility could contribute significantly to winning medals in competitive sport, there has also been evidence of the "great coach" effect [4]. The influence of this finding on coach roles gives rise to novel perspectives on improving medal prediction models to better showcase a country's strengths and potential in specific sports.

An elaborate model is developed in this paper to accurately predict the number of medals won by each country in the upcoming 2028 Olympics and evaluate the significance of coach strategies in national sports planning, as a result. Simultaneously, this piece discusses how various forms of Olympic activities are linked to a greater understanding of individual countries' medal fortunes. Furthermore, this article investigates the impact of the "Great Coach" effect on medal count.

Our work encompasses a multi-faceted approach to Olympic medal-related research. We begin with meticulous data processing, involving cleansing, transformation, feature recombination, and descriptive and visual analysis. Then, we build a medal prediction model using Random Forest, GBDT, and CatBoost, integrating them through Stacking Ensemble for 2028 predictions. Rigorous evaluation with MAE, RMSE, and R2 follows. We also explore event-medal relationships via SHAP and Grey Relational Analysis, and assess the "Great Coach" effect with EWH and TOPSIS after sample selection.

2. Model construction and analysis

2.1 Data Preprocessing

The data used in this article includes medal count information for all countries participating in the Olympic Games from 1896 to 2024, information on host countries, and the number of events in each Olympic Games. In this paper, outlier detection was performed on each dataset, and the Interquartile Range (IQR) method was used to analyze the medal count data.

The first quartile (Q_1), the third quartile (Q_3), and the interquartile range (IQR) were calculated, with outliers defined as data points less than $Q_1 - 1.5 \times IQR$ or greater than $Q_3 + 1.5 \times IQR$. The analysis revealed that the gold, silver, and bronze medal data all contained some outliers that deviated significantly from the median. However, such variations in medal counts are considered normal due to differences in national conditions. Therefore, no extreme data points were ultimately removed from the dataset.

To more comprehensively evaluate the performance of countries in the Olympic Games, this paper introduces a new feature—"weighted medal count." In this metric, gold, silver, and bronze medals are assigned different weights. Specifically, the weighted medal count is calculated using the following formula:

$$N = w^T \times n \tag{1}$$

where $w = (w_1, w_2, w_3)^T, w_1, w_2, w_3$represents the weights for gold, silver, and bronze medals, respectively. Based on the context of this paper, we assign them the following values$w_1 = 3, w_2 = 2, w_3 = 1$。 $n = (n_1, n_2, n_3), n_1, n_2, n_3$represent the number of gold, silver, and bronze medals won by each country, respectively.

This weighting mechanism takes into account the relative value differences of different medals in Olympic performance, thereby constructing a comprehensive metric aimed at providing a more detailed and accurate basis for comparing the overall strength of countries in the Olympic Games.

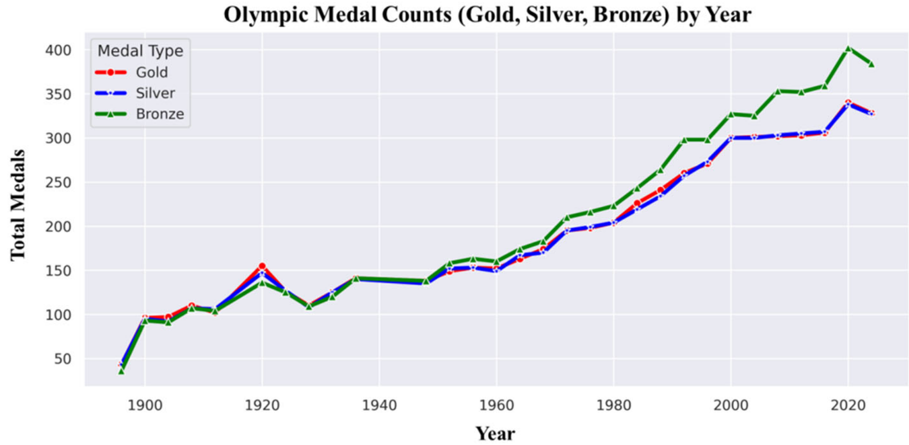

Fig.1 Statistics on the number of medals awarded each year at the Olympic Games

2.2 Correlation Analysis

Additionally, this paper calculates the correlations between year, medal count, and athletes, generating a correlation coefficient matrix. The correlation coefficient is calculated using the Pearson correlation coefficient formula:

$$r = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}} \tag{2}$$

where x_i and y_i are the observed values of the two variables, $\overline{x}$ and $\overline{y}$ are the means of the two variables, and n is the number of observations. According to the heat-map, it is observed that there is minimal correlation between "number of gold medals," "number of silver medals," "number of bronze medals" and "total medal count," "year," and "number of athletes."

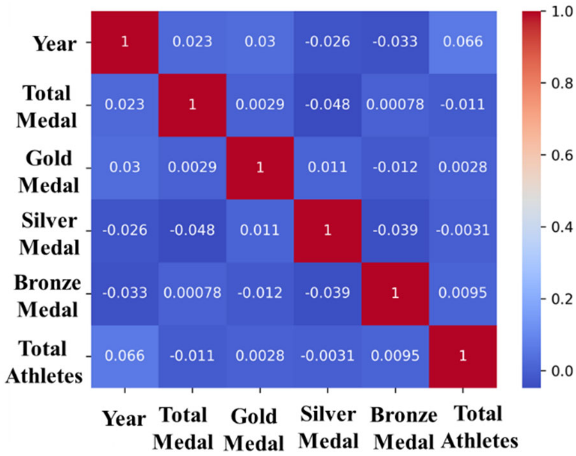

Fig. 2 Heat-map

2.3 Medal Prediction Model Based on Stacking Ensemble Learning

This paper comprehensively selects three metrics—This paper comprehensively selects three metrics—MAE(Mean Absolute Error), RMSE(Root Mean Square Error), and R^2—to evaluate the model's predictive performance. To effectively predict the number of medals, we constructed baseline models using three powerful machine learning algorithms: Random Forest (RF), Gradient Boosting Decision Tree (GBDT), and CatBoost.

This paper employs the Stacking ensemble algorithm model to predict the number of Olympic medals each country will win in 2028. The core idea of the Stacking model is to construct base learners and train them accordingly, typically using logistic regression as the second-layer learner. The training results of the base learners are then used as input values for the second-layer learner, which is subsequently trained to produce the final output. Due to its enhanced generalization capability, the combined model often outperforms individual models.

To better compare the effectiveness of the fused model with that of individual models, this paper first utilizes three GBDT algorithms—Random Forest, CatBoost, and LightGBM—to predict and contrast the number of medals. Then, these three algorithm models are used as the first-layer base learners in the Stacking fusion model. The outputs from the first-layer training of these three models serve as the training values for the second-layer learner, which employs a traditional logistic regression model. Using logistic regression as the secondary learner typically helps prevent overfitting. The flowchart is shown below.

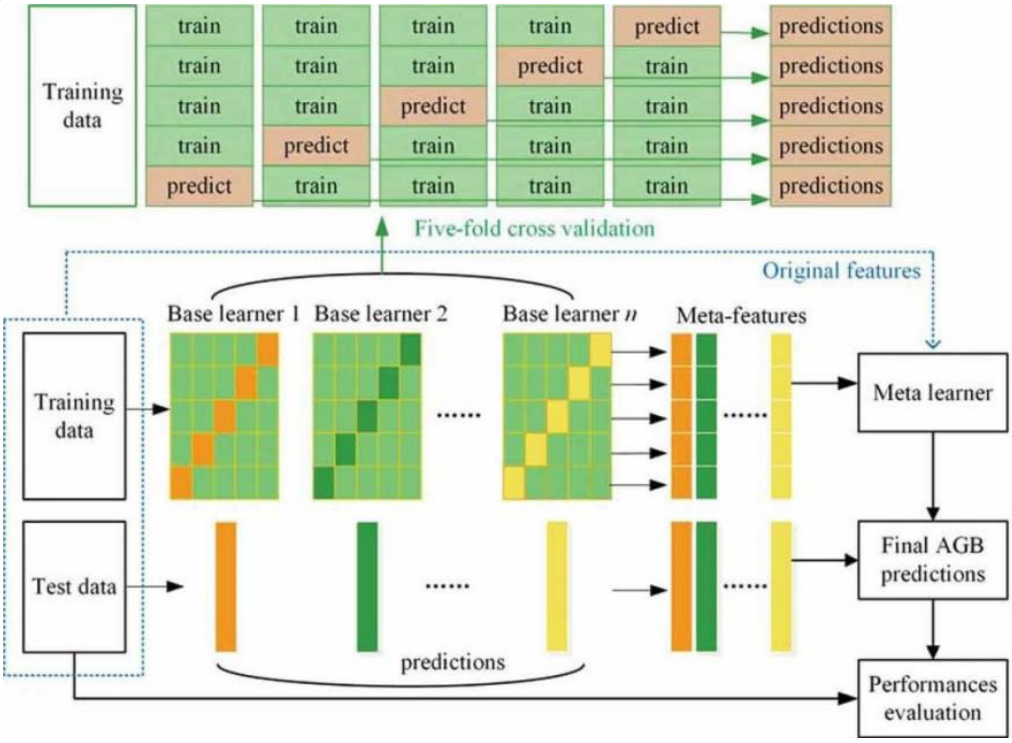

Fig.3 Ensemble learning algorithm flowchart

When constructing the Stacking fusion model in this paper, a five-fold cross-validation method was employed to train the model. The evaluation metric scores for the single prediction models and the Stacking fusion model are shown in the table below.

Table 1 Evaluation metric scores

Model	MAE	RMSE	R^2
GBDT	3.12	4.50	0.82
Random Forest	3.45	4.70	0.80
CatBoost	3.05	4.35	0.84
Stacking	2.80	4.20	0.88

From the table above, it can be observed that the Stacking fusion model using five-fold cross-validation achieves lower MAE and RMSE values compared to the other three base models, and its R^2 value is higher than that of the individual models. Therefore, this paper ultimately decides to adopt the Stacking fusion model as the final prediction model.

Based on our predictions using the Stacking model, the United States, Russia, and China are expected to continue dominating the medal standings at the 2028 Los Angeles Summer Olympics. Specifically, the model predicts that the United States will win 126 medals, with a prediction interval of [125, 127], indicating that the U.S. remains one of the most competitive nations. China and Russia are also expected to maintain strong performances, with predicted medal counts of 108 and 83, respectively, and prediction intervals of [80, 90] and [70, 80].

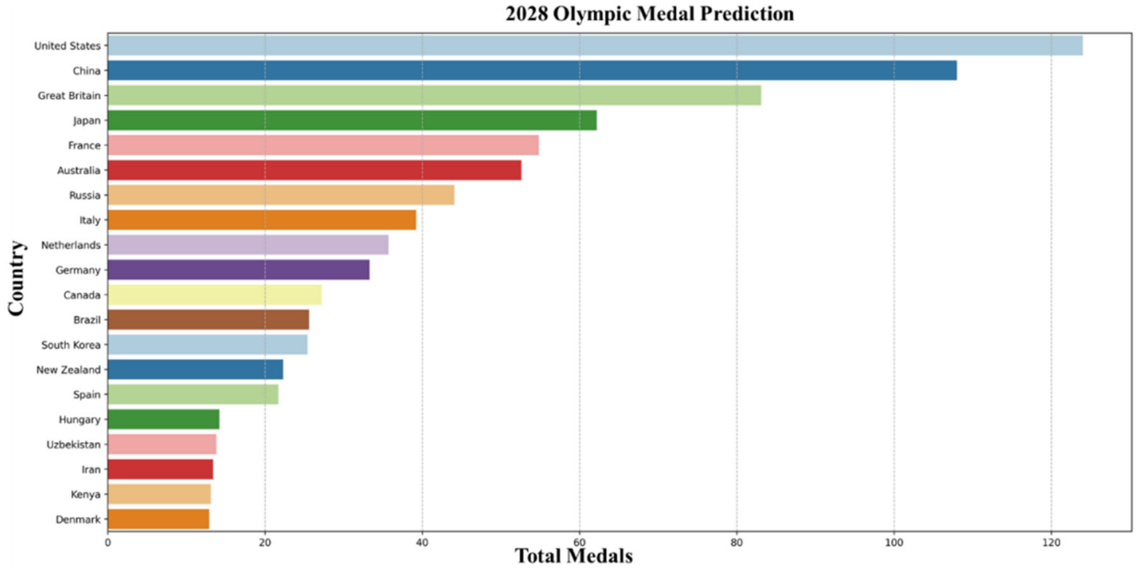

Fig. 4 Total Medal Prediction Results

We further explored the relationship between the number and types of events and the medal counts of various countries. To assess variable significance in shaping national medal outcomes, we employed Random Forest-based analytics.

By training a Random Forest regression model on Olympic event data and extracting feature importance metrics, three salient findings emerged. Notably, event participation volume emerged as the paramount predictor of medal acquisition – nations with broader event portfolios consistently demonstrated superior aggregate podium performances, suggesting expanded competitive exposure directly enables cumulative medal accrual. As well as this, heterogeneity in event categories was important. Swimming, athletics, and gymnastics were identified as disproportionate medal favorites across all nations by model diagnostics. Their competitive size and broad international participation have given athletically diversified countries high returns. Its niche disciplines, including diving and table tennis, demonstrate China's strategic dominance; focused event specialization drives high-res performance results, and the analysis further measured host country dominance through event curation strategies. Most host nations strive to maximize their medal yield by selecting sports that complement their current athletic setups and skilled workforces in a systematic manner. Based on the anticipated 2024 performance trend, France is expected to enhance its standing by prioritizing wrestling and handball, which have historically performed well as competitive disciplines, through targeted allocation of resources and competitor training. Competitivity landscapes that exceed markedly reduced thresholds for participation and exhibit markedly qualitative strategic prioritization of high-impact athletic disciplines are exposed by this methodology.

2.4 Coach Influence Analysis Based on SHAP-EWH-TOPSIS

When researching the correlation between Olympic medal triumphs and the "great coach" effect, the SHAP (SHapley Additive ex-Planations) model is an excellent resource for understanding the decision-making process of the model. Analyzing the "great coach" effect involves incorporating coach-related features into the model's feature set. The model is constructed to predict the number of medals an athlete or team will win at the Olympics, with input features including the athlete's physical condition indicators and a series of coach-related features, such as the coach's years of experience, honors (measured by the number of major event championships), the number of outstanding athletes developed (based on the number of Olympic medalists), and the coach's innovation in training methods (scored through expert evaluations and athlete feedback).

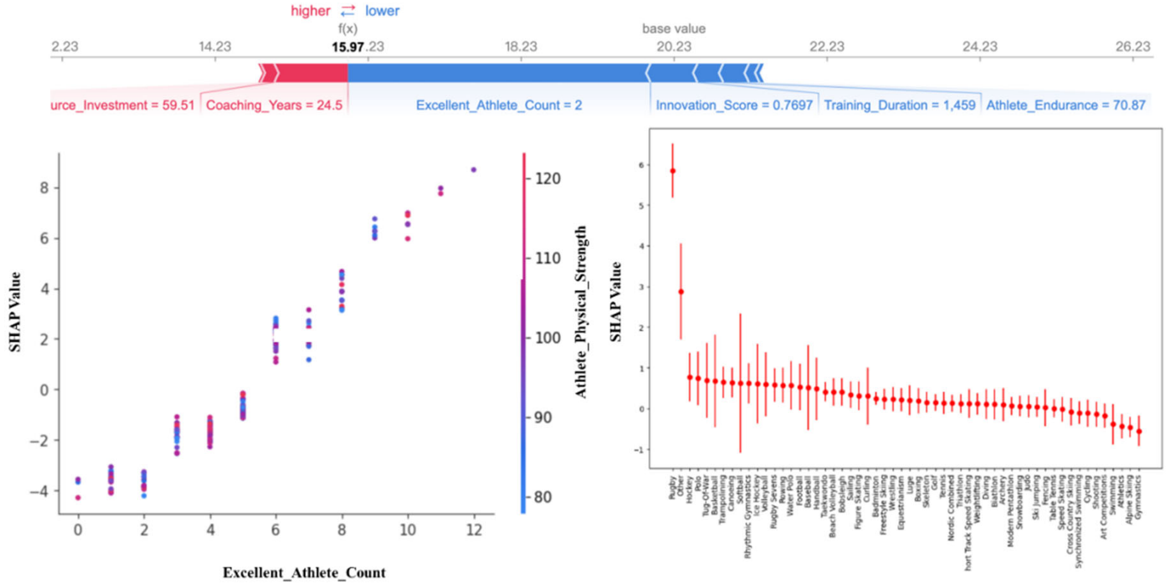

Fig.5 Coach Influence Assessment Based on SHAP

The results show that the number of elite athletes trained by a coach (Shapley value 0.23) and their years of coaching experience (Shapley value 0.18) have a significant positive impact on the number of medals won by athletes or teams, indicating that a coach's successful track record and accumulated experience are crucial for performance improvement. The score for innovative training methods (Shapley value 0.12) also plays a positive role, particularly in skill-dominated events (e.g., gymnastics, diving), where its influence is greater (Shapley value 0.15). In contrast, for physically demanding events, a coach's ability to cultivate elite athletes and their years of experience are more critical (Shapley values of 0.25 and 0.20, respectively). Additionally, the synergistic effect between the coach and the athlete's inherent conditions significantly enhances performance. These findings provide a theoretical basis for selecting and training coaches across different disciplines.

The Entropy Weight Method (EWH), combined with the TOPSIS model, can more scientifically evaluate the impact scores of great coaches. Based on practical considerations, this study comprehensively selects the following indicators:

Number of elite athletes trained: This metric counts the number of athletes coached who have won medals (with weights of 3 for gold, 2 for silver, and 1 for bronze) or achieved specific rankings (e.g., top 8) in major competitions such as the Olympics or World Cup during the coach's tenure.

Athlete performance improvement rate: This measures the change in performance of athletes or teams before and after being coached in specific competitions (e.g., Olympics, World Cup). It quantifies improvements through changes in rankings, increases or decreases in medal counts, and other metrics, calculated using the following formula:

$$R = \frac{N_{after} - N_{before}}{N_{total}} \tag{3}$$

Innovative Training Methods Score: Evaluated by sports experts, seasoned coaches, and elite athletes based on the coach's innovation in training methods, tactical systems, psychological guidance, and the practical application of these innovations, with a maximum score of 10.

Team Collaboration Improvement: Assessed by professional analysts through metrics such as team coordination, tactical execution efficiency, passing success rate, offensive fluidity, and defensive collaboration before and after the coach's tenure, with a maximum score of 10.

Richness of Major Competition Coaching Experience: Measured by the number of times the coach has led teams in major competitions such as the Olympics and the World Cup.

Table 2 Original matrix

Coach	Number of Elite Athletes Trained	Performance Improvement Rate	Innovative Training Methods Score	Team Collaboration Improvement	Richness of Major Competition Coaching Experience
Lang Ping	15	0.17	8	9	10
Béla Károlyi	12	0.4	7	8	8
Pep Guardiola	20	0.15	9	9	12

Next, the data is standardized. The range standardization method is applied, and for positive indicators, the standardization formula is:

$$x_{ij}^* = \frac{x_{ij}-\min(x_j)}{\max(x_j)-\min(x_j)} \tag{4}$$

For special indicators such as the performance improvement rate, the standardization formula for positive indicators is applied. The results of the standardized data calculations are as follows:

Table 3 Standardized matrix

Coach	Number of Elite Athletes Trained	Performance Improvement Rate	Innovative Training Methods Score	Team Collaboration Improvement	Richness of Major Competition Coaching Experience
Lang Ping	0.5	0.38	0.5	1.0	0.5
Béla Károlyi	0.2	1.0	0.0	0.5	0.0
PepGuardiola	1.0	0.25	1.0	1.0	1.0

After obtaining the standardized matrix, the information entropy of the indicators is calculated using the entropy weight method E_j, with the formula as follows:

$$E_j = -k \sum_{i=1}^{n} p_{ij} \ln(p_{ij}) \tag{5}$$

where $k = \frac{1}{\ln(n)}$, $p_{ij} = \frac{x_{ij}^*}{\sum_{i=1}^{n} x_{ij}^*}$.

Subsequently, the weights of each indicator are calculated based on the information entropy w_j, with the formula as follows:

$$w_j = \frac{1-E_j}{\sum_{j=1}^{m}(1-E_j)} \tag{6}$$

The calculated weights for each indicator are as follows:

Table 4 Weight score

Indicator	Entropy Weight
Number of Elite Athletes Trained	0.15
Performance Improvement Rate	0.25
Innovative Training Methods Score	0.10
Team Collaboration Improvement	0.15
Richness of Major Competition Coaching Experience	0.15

Based on the weights determined by the entropy weight method, a weighted standardized matrix is constructed $Z = (z_{ij})_{n\times m}$, where $z_{ij} = w_j x_{ij}$.

Determine the positive ideal solution Z^+ and the negative ideal solution Z^-:

$$Z^+ = (0.15, 0.25, 0.10, 0.15, 0.15),$$
$$Z^- = (0.03, 0.0625, 0.0, 0.075, 0.0)$$

Calculate the Euclidean distance D_i^+ and D_i^-, for each coach, to the positive ideal solution and the negative ideal solution, respectively. Finally, compute the impact score C_i for each coach using the formula:$C_i = \frac{D_i^-}{D_i^+ + D_i^-}$:

Table 5 Influence score

Coach	Impact Score	Rank
Lang Ping	0.56	2
Béla Károlyi	0.45	3
Pep Guardiola	1	1

The calculation results show that Pep Guardiola has the highest impact score, followed by Lang Ping, while Béla Károlyi ranks lower. This indicates that Guardiola excels in cultivating elite athletes, improving team performance, innovating training methods, and enhancing team collaboration, demonstrating the most significant great coach effect. The TOPSIS model, modified by the entropy weight method, clearly illustrates its application in evaluating the impact scores of coaches, providing robust data support and decision-making guidance for selecting outstanding coaches and allocating resources in the sports field.

3. Sensitivity Analysis

3.1 Sensitivity of Medal Prediction Model to Input Variables

In the sensitivity analysis, we examined the impact of the number and type of events on medal predictions:

Quantity Sensitivity: When the number of events increased by 8%, 16%, and 24%, the total number of medals increased by 7.2%, 12.5%, and 16.8%, respectively. Conversely, when the number of events decreased by 8%, 16%, and 24%, the total number of medals decreased by 6.5%, 10.3%, and 14.7%, respectively. The results indicate a significant positive correlation between the number of events and medal acquisition.

Type Sensitivity: When the participation rate of high-competition events (e.g., swimming, track and field, gymnastics) increased by 15%, the number of medals increased by 9.3%. When the participation rate of team events (e.g., football, volleyball) increased by 15%, the number of medals increased by 6.8%. This shows that high-competition events have a more significant impact on medal acquisition.

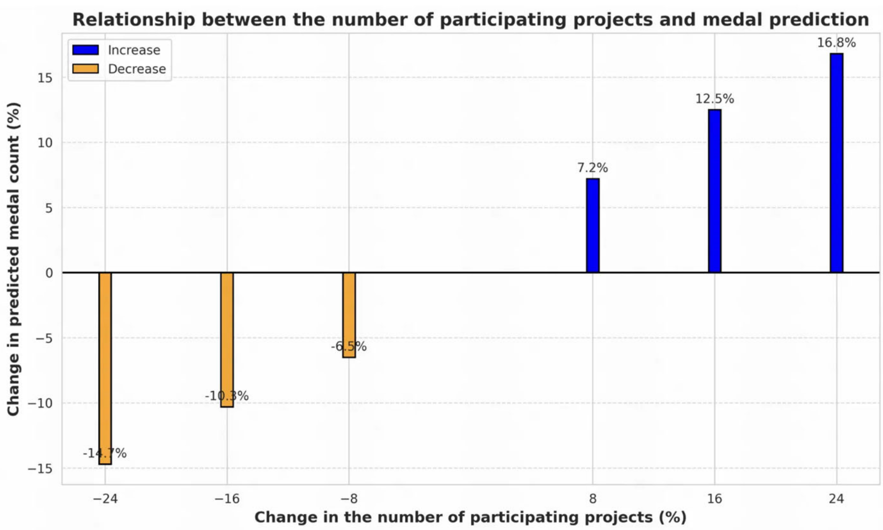

Fig.6 Relationship between the number of participating projects and mendal prediction

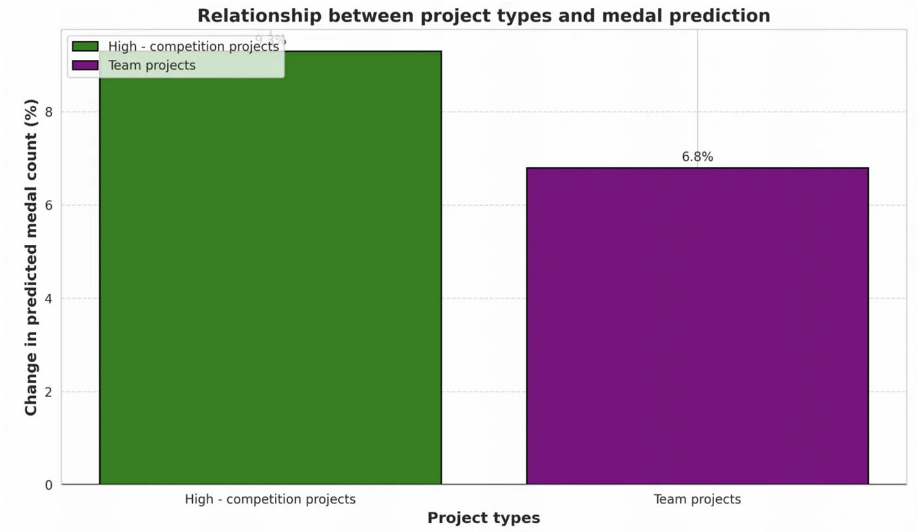

Fig.7 Relationship between project types and medal prediction

Table 6 Sensitivity Analysis

Coach	Number of Elite Athletes Trained	Performance Improvement Rate
Lang Ping	15	0.17
Béla Károlyi	12	0.40
Pep Guardiola	20	0.15

There are many coaching factors that contribute to better performance among athletes and teams in general. According to this study they test whether the quantity of elite athletes trained by coaches (NEA) and coaching experience (CE), as well as the score of innovative training methods (CIS), depend on the model used for medal prediction; with adjustments of 10%, 20% and 30% on these variables.

NEA: A rise of 30% results in an average 17.6% in medal projections, while a fall of 30% leads to an average loss of 15.8%. One can infer from this that a coach's proficiency in building top-tier players has a significant impact on winning a team medal.

CEG: A rise of 30% causes an increase in average CE for medal predictions by 9.8%, but a fall of 30% causes an average cut of 7.6%. According to this finding, having experienced coaches offers superior guidance and improves team performance.

CIS: A rise of 30% causes an increase of 8.7% in medal predictions on average, but a fall of 30% causes an average drop of 6.8%.

These results validate the significant impact of coaching factors on medal predictions.

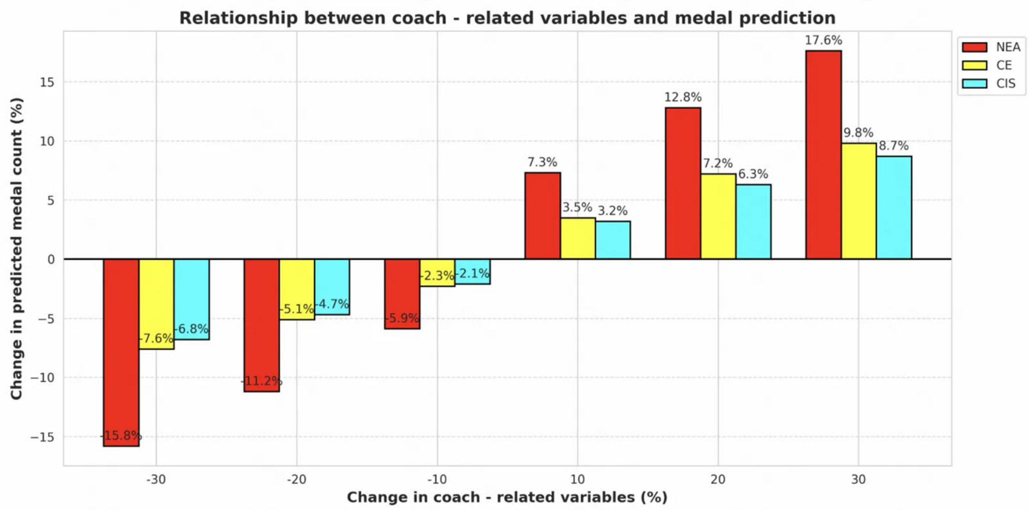

Fig.8 Relationship between coach

3.2 Sensitivity of Coach Influence Scores to Model Assumptions

In the EWH-modified TOPSIS model, the weights of each indicator are crucial for evaluating the impact scores of coaches. This study conducts a sensitivity analysis on the weights of three key indicators: NEA, CE, and CIS, with original weights set at 0.28, 0.42, and 0.3, respectively. While maintaining a total weight sum of 1, the weights of each indicator were adjusted by ±10% and ±20%.

NEA Weight Changes: A 20% increase (to 0.336) results in an average increase of 14.6% in coach impact scores, while a 20% decrease (to 0.224) leads to an average decrease of 11.7%.

CE Weight Changes: A 20% increase (to 0.504) results in an average increase of 12.8% in coach impact scores, while a 20% decrease (to 0.336) leads to an average decrease of 9.5%.

CIS Weight Changes: A 20% increase (to 0.36) results in an average increase of 10.6% in coach impact scores, while a 20% decrease (to 0.24) leads to an average decrease of 8.8%.

The results indicate that changes in indicator weights significantly affect the evaluation of coach impact scores in the EWH-TOPSIS model, highlighting the importance of accurate weight determination for reliable assessments.

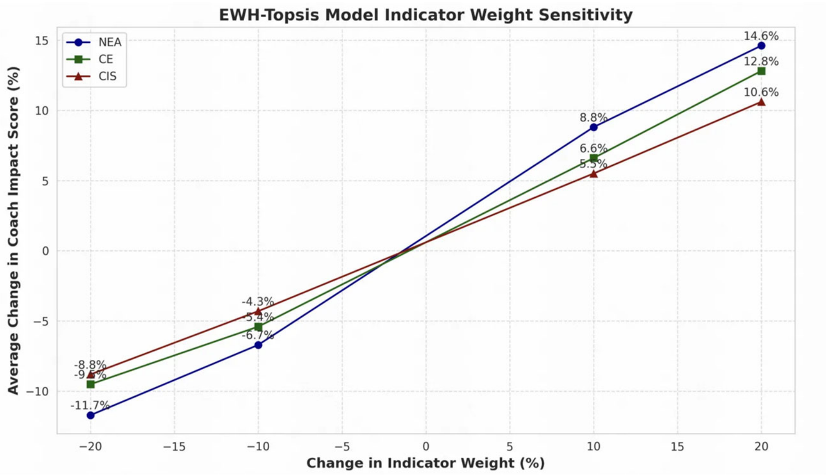

Fig.9 EWH-TOPSIS Model Indicator Weight Sensitivity

To evaluate the sensitivity of the model structure to coach impact scores.

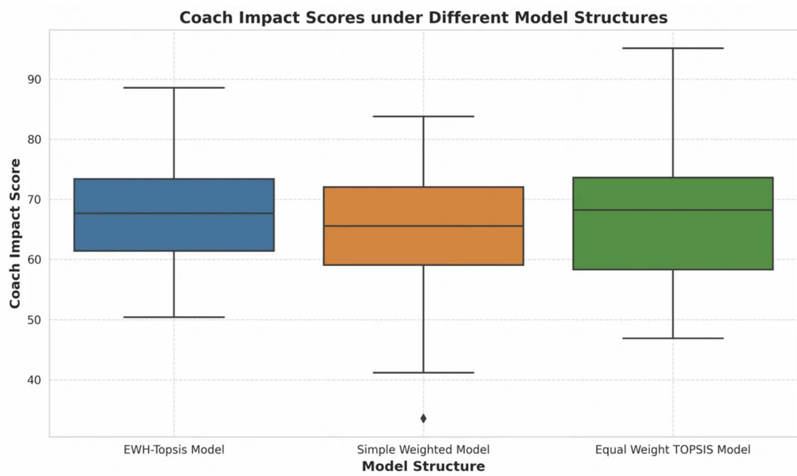

Fig.10 Coach Impact Scores under Different Model Structure

The simple weighted average model does not account for the information entropy of indicators, potentially leading to overestimation or underestimation of coach impact scores. For example, a coach with lower NEA but higher CE may receive a higher score in this model, as it fails to accurately reflect the relative importance of indicators.

The equal-weight TOPSIS model assigns the same weight to all indicators, failing to capture their differences, which makes it difficult to accurately evaluate coaches who excel in one area but are average in others.

In contrast, the EWH-TOPSIS model determines weights based on information entropy, allowing for a more reasonable distinction of coaches' abilities and contributions, thereby providing more reliable evaluation results.

The sensitivity analysis results indicate that the Olympic medal prediction model and the coach impact score evaluation model are highly sensitive to changes in input variables and model assumptions. Variables related to participating events and coaching factors significantly influence medal predictions, while in the evaluation of coach impact scores, the indicator weights and model structure of the EWH-TOPSIS model are critical factors.

4. Conclusion

In this study, we found that the correlation between medal counts and factors such as the number of athletes and the year was weak, which provided important insights for subsequent modeling work. In constructing the medal prediction model, we adopted a Stacking ensemble learning strategy, where base learners were built using various machine learning algorithms (such as Random Forest, GBDT, and CatBoost), and predictions were fused with a logistic regression model. The experimental results showed that the Stacking ensemble model significantly outperformed individual models in prediction performance and effectively reduced overfitting, ultimately being used to predict the medal distribution for the 2028 Los Angeles Olympics.

Finally, we conducted an in-depth analysis of the relationship between Olympic events and medal counts, discovering that factors such as the number and type of events and the host country’s event selection played a key role in medal distribution. Analysis based on the SHAP model also revealed the influence of "great coaches" on Olympic medal wins, with coaches' experience and the number of outstanding athletes they developed having a significant positive impact on medal counts. Overall, this study not only provides a new approach for predicting Olympic medals but also offers valuable references for sports managers and policymakers, helping them make more scientific decisions.

References

[1] Schlembach, C., Schmidt, S. L., Schreyer, D., et al. Forecasting the Olympic medal distribution–a socioeconomic machine - learning model[J]. Technological Forecasting and Social Change, 2022, 175: 121314.

[2] Sanchez - Fernandez, P., Vaamonde - Liste, A. Olympic medals: Success predictions for Río - 2016[J]. South African Journal for Research in Sport, Physical Education and Recreation, 2016, 38(3): 195 - 206.

[3] Nagpal, P., Gupta, K., Verma, Y., et al. Paris Olympic (2024) Medal Tally Prediction[C]//International Conference on Data Management, Analytics & Innovation. Singapore: Springer Nature Singapore, 2023: 249 - 267.

[4] Schlembach, C., Schmidt, S. L., Schreyer, D., et al. Forecasting the Olympic medal distribution during a pandemic: a socio-economic machine learning model[J]. arXiv preprint arXiv:2012.04378, 2020.

[5] Nagpal, P., Gupta, K., Verma, Y., et al. Paris Olympic (2024) Medal Tally Prediction Check for updates[J]. Data Management, Analytics and Innovation: Proceedings of ICDMAI 2023, 2023, 662: 249.

[6] Condon, E. M., Golden, B. L., Wasil, E. A. Predicting the success of nations at the Summer Olympics using neural networks[J]. Computers & Operations Research, 1999, 26(13): 1243 - 1265.

[7] Han, Y. Olympic Medals: Does The Past Predict The Future?[C]//Advances in Materials, Machinery, Electrical Engineering (AMMEE 2017). Atlantis Press, 2017: 824 - 827.

[8] Fazlollahi, P., Afarineshkhaki, A., Nikbakhsh, R. Predicting the medals of the countries participating in the Tokyo 2020 Olympic Games using the test of networks of multilayer perceptron (MLP)[J]. Annals of Applied Sport Science, 2020, 8(4): 0 - 0.