International Journal of Advanced Science

**Research article**

https://doi.org/10.70731/0dr6zf77

# 2028 Olympic Medal Prediction Model: Integrating Random Forest and Monte Carlo Simulation

**Qixuan Zhang** [a,*], **Wanying Chen** [a], **Zhefan Chen** [a], **Chengtian Liang** [a]

[a] *School of Physics, Hangzhou Normal University, Hangzhou 311121, China*

KEYWORDS

*Olympics;*
*Medal;*
*Prediction;*
*Random Forest;*
*Monte Carlo simulation;*
*Linear Regression*

ABSTRACT

This article investigates the prediction of medal outcomes for the 2028 Los Angeles Summer Olympics and establishes a multi-model framework to analyze and forecast the potential results for various countries. A random forest regression model is used to analyze historical medal counts and competition data, alongside a Monte Carlo simulation to quantify uncertainty. The findings suggest that the United States and China are expected to lead in both the number of gold medals and total medals. Additionally, an analysis predicts that two countries may secure medals for the first time in 2028, with the associated error quantified. A weighted scoring model evaluates the importance of different Olympic events for various countries, revealing that host countries tend to excel in sports with historical advantages or strategic investments. The impact of the "Great Coach" effect on medal performance is also examined, demonstrating that high-profile coaches significantly improve the performance of national teams. Finally, the study compares the differences in medal distributions among countries and highlights notable examples of athletes whose performances have evolved over time.

## INTRODUCTION

The 2024 Paris Summer Olympics is a global sports event, with the United States ranking first with 126 medals, and China tied with the United States in gold medals, winning 40 gold medals, demonstrating its strong strength. As the host country, France ranks fifth with 16 gold medals and fourth in the total number of medals, highlighting its local advantage. The UK ranks seventh with 14 gold medals and third in total medal count.

As the Olympic Games approach, medal predictions have become the focus, with historical data providing the foundation for predictions. The announcement of participating athletes greatly improves the accuracy of predictions. The number and type of Olympic events have a profound impact on medal allocation, with the advantages of traditional strong countries and the selection of events by the host country playing an important role in the medal table.

In addition to the dominance of traditional powers, countries such as Albania, Cape Verde, the Dominican Republic, and Saint Lucia have made their debut on the medal table. More than 60 countries have not yet won

* Corresponding author. E-mail address: 1523089730@qq.com

Received 2 March 2025; Received in revised from 5 September 2025; Accepted 28 October 2025; Published online 30 November 2025.

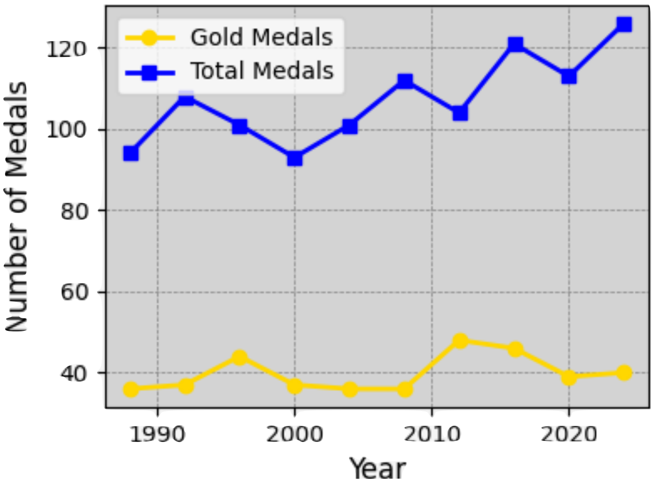


**Figure 1 | Top 10 Countries in 2024 Olympics**

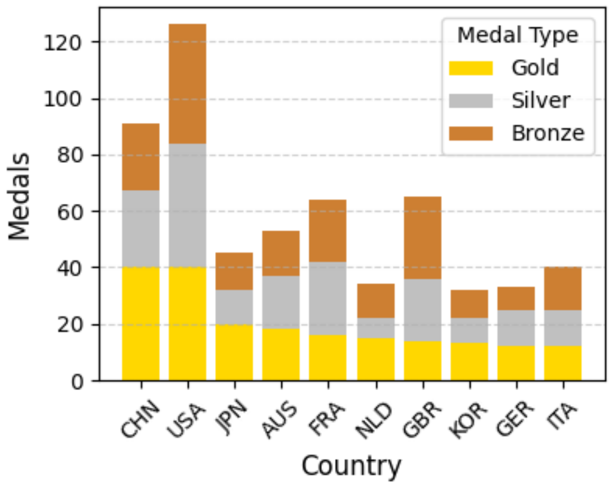


**Figure 2 | USA Medal Distribution in 10th Olympics**

awards, and the global sports landscape is still full of uncertainty and opportunities.

The influence of the coach cannot be ignored. The flexibility of coaches in cross-border coaching, such as the success of Lang Ping and Bella Karoli, demonstrates the "great coaching effect" and significantly improves team performance.

## ASSUMPTIONS AND NOTATIONS

- There is a predictable relationship between the number of medals and the year and country. Supposing that the model can capture these relationships and provide effective evidence for future medal predictions.
- Supposing that all historical medal data (including the number of gold, silver, and bronze medals) is complete and accurate, and can provide meaningful basis for future medal predictions.
- Supposing that by 2028, the number of countries participating in the Olympics will not undergo drastic changes, ignoring the impact of large-scale national disintegration or the addition of new countries.
- Supposing that all participating countries participate fairly and justly, without any large-scale fouls (such as doping)
- Supposing that the level of most participating countries will not undergo a sudden change, ignoring a few upsets
- Supposing that future Olympic Games will not be affected by significant rule changes or external events such as politics or natural disasters.

we use some symbols for constructing the model as **Table 1.**

## DATA PROCESSING

We obtained the original dataset from three files: summerOly_medal_counts, summerOly_athletes, and summerOly_hosts, which include data on Olympic organizers, athletes' awards, and medal counts of participating countries.

During data cleaning, we identified outliers. We retained data from countries that participated in the 2024, 2020, and 2016 Olympics, excluding non-existent countries (e.g., the Soviet Union) and unstable entities (e.g., the CIS).

We found that there are missing project data for each year. When dealing with missing values, we used queries to supplement relevant information to ensure data integrity.

Additionally, we extracted medal information for each participating country across different events and marked the host country for each year.

## MODELING

### Prediction of Olympic Medals and First-Time Medalists via Multi-Model Framework

We use a random forest regression model to predict the medal table for the 2028 Summer Olympics in Los Angeles, USA based on historical data of Olympic medal wins. We not only predicted the number of gold, silver, and bronze medals for each country, but also quantified the prediction interval for the number of medals for each country through Monte Carlo simulation and evaluated the uncertainty. The model results also analyzed which countries may progress or regress in 2028, helping to determine which countries are most likely to break through or regress.

**Table 1 | Symbols of the model**

| Symbol | Description |
|---|---|
| $Y$ | Target variable matrix, representing the number of medals (gold, silver, bronze) |
| $X$ | Feature matrix, containing features of year and country (one-hot encoded) |
| $N$ | Number of countries (for simulation and statistics) |
| $T$ | Total number of simulations |
| $G$ | Number of Gold Medals (number of gold medals per country) |
| $S$ | Number of Silver Medals (number of silver medals per country) |
| $B$ | Number of Bronze Medals (number of bronze medals per country) |
| $g$ | Predicted number of gold medals (predicted number of gold medals for a single country) |
| $s$ | Predicted number of silver medals (predicted number of silver medals for a single country) |
| $b$ | Predicted number of bronze medals (predicted number of bronze medals for a single country) |
| $\Delta G$ | Change in the number of gold medals (difference between predicted and actual number of gold medals) |
| $\Delta S$ | Change in the number of silver medals (difference between predicted and actual number of silver medals) |
| $P$ | Total predicted number of medals (gold, silver, bronze) |
| $p$ | Prediction of the number of medals for the i-th country |
| $P_{sin}$ | Predicted total number of medals from Monte Carlo simulation results |
| $\mu$ | Average value (expected value) in simulation |
| $\sigma$ | Standard deviation in simulation |
| $f_{model}$ | Prediction function of regression model |
| $\hat{P}$ | Number of medals predicted by the model (standardized prediction results) |
| $e$ | Simulation error or random error |

The mathematical model we define is as follows:

$$f_{\text{model}}(X) = \sum_{i=1}^{T} \frac{1}{T} \left(h_i(X)\right) \quad (1)$$

The second part predicts which countries that have not yet won medals may win their first medals in 2028. These countries usually perform weaker, but with improved athlete training and an increase in new events, they are expected to break through their historical performance. We predict which countries are likely to break through and win medals based on historical data and national performance trends.

To achieve this goal, we have generated forecast values for each country, indicating their likelihood of winning their first medal in 2028. This problem is transformed into a binary classification problem, where we use a random forest regression model to predict the probability of a country's success.

For each country , we define a binary variable $Y_i$, where:

$$Y_i = \begin{cases} 1, & \text{if country } i \text{ wins its first medal} \\ 0, & \text{if country } i \text{ does not win its first medal} \end{cases} \quad (2)$$

Use a random forest regression model to predict the probability $P(Y_i = 1 \mid X_i)$, which is the probability that country i will win its first medal in the 2028 Olympics, based on the features $X_1, X_2, \ldots, X_n$, where $X_1, X_2, \ldots, X_n$ are the characteristics of country $i$.

The regression model is as follows:

$$P(Y_i = 1 \mid X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}} \quad (3)$$

where $\beta_0, \beta_1, \ldots, \beta_n$ are the regression coefficients, and $X_1, X_2, \ldots, X_n$ are the features used to predict whether country i will win its first medal.

On this basis, we adopted the Monte Carlo simulation method, which uses repeated random sampling for numerical calculations to help us quantify the uncertainty of predictions.

By repeatedly sampling the possible values of $P(Y_i = 1)$, we can simulate different scenarios and estimate the uncertainty.For example, if the predicted probability of a country winning a medal is 0.25, and we assume a normal distribution, we can use Monte Carlo methods to generate possible values of the probability

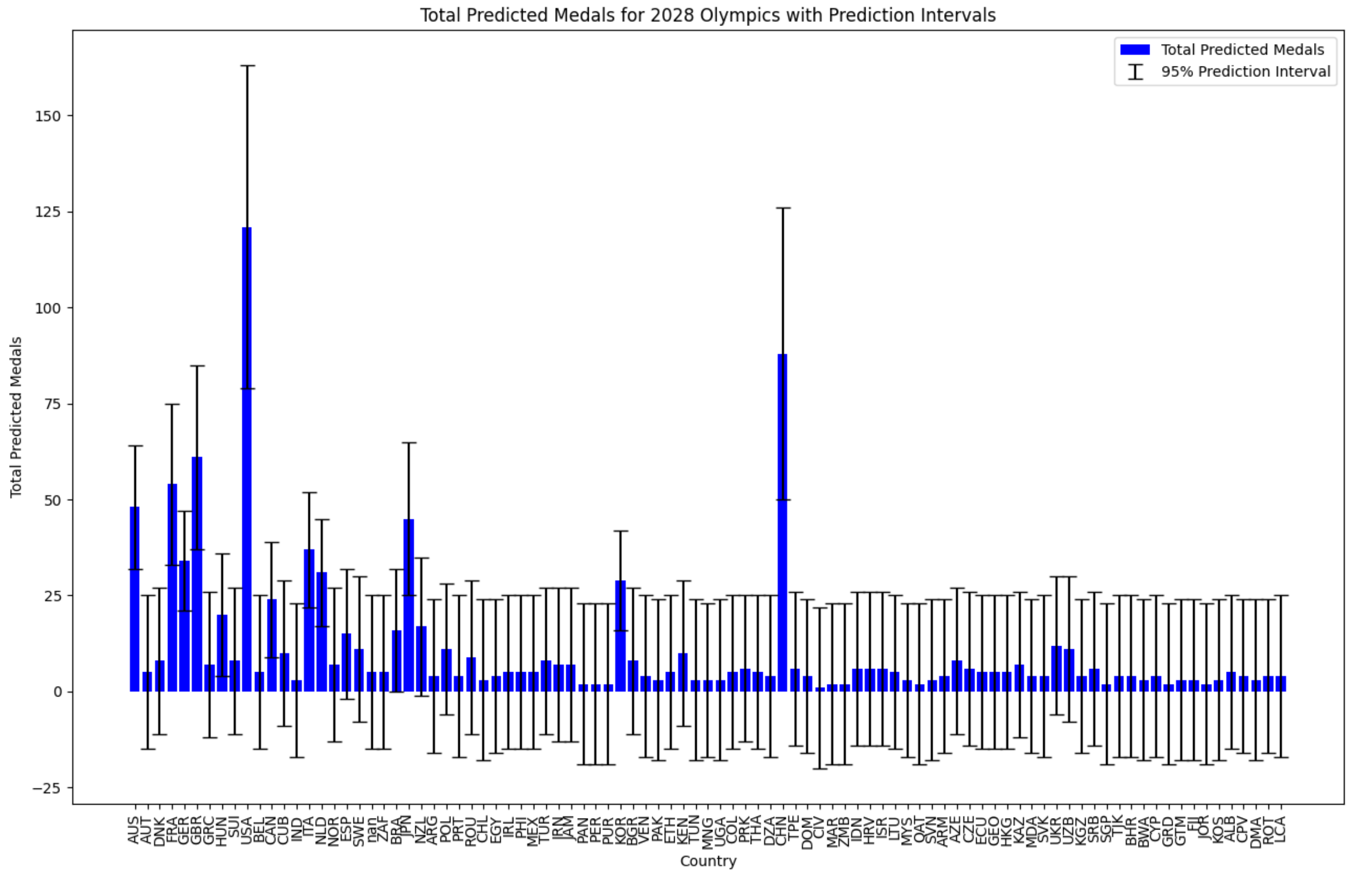


**Figure 3 | Predict the Number of Medals with a 95% Confidence Interval**

and estimate the likelihood of the country winning its first medal.

$$P(Y_i = 1) \sim N(\hat{P}, \sigma) \quad (4)$$

Where $\hat{P}$ is the predicted probability, and $\sigma$ is the standard deviation, based on the historical data used to calculate this probability.

In the analysis process, we paid special attention to the relationship between the total number of medals and the number of winning countries, and further speculated through simulations which countries are most likely to win medals for the first time in 2028.

Based on our trained random forest model, the 2028 Olympic medal predictions show that the United States and China will continue to lead in both gold and total medals. Monte Carlo simulations provide more accurate prediction intervals and assess the fluctuation range for each country's medal count. It also quantifies prediction uncertainty by modeling various scenarios. The narrower intervals for the United States and China indicate higher accuracy, while the wider intervals for other countries reflect greater uncertainty.

As shown in the **Figure 3**, it reflects the predicted number of medals and a 95% confidence interval. The blue bar represents the predicted number of medals, and the black line segment represents the confidence interval.

Based on our trained random forest model, the 2028 Olympic medal predictions show that the United States and China will continue to lead in both gold and total medals. Monte Carlo simulations provide more accurate prediction intervals and assess the fluctuation range for each country's medal count.It also quantifies prediction uncertainty by modeling various scenarios. The narrower intervals for the United States and China indicate higher accuracy, while the wider intervals for other countries reflect greater uncertainty.

The United States is expected to win between 38 and 42 gold medals, with silver and bronze ranges from 36 to 40 and 32 to 36, respectively. China is forecasted to earn between 36 and 40 gold medals, with silver and bronze ranges from 30 to 34 and 28 to 32. These results suggest both countries are likely to secure the top two positions in gold and total medal counts.

Based on model predictions and analysis of athlete strength, competition participation, and coach system optimization, we anticipate the following countries may surpass their 2024 performance in 2028.

India, with the rise of emerging sports and improved athlete training, is expected to win more medals, particularly in badminton and wrestling. Russia, despite polit-

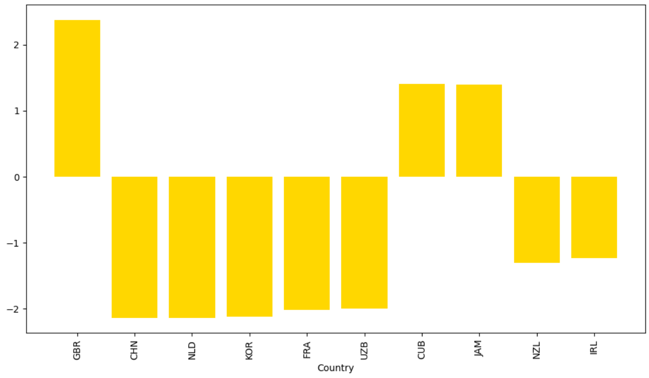


**Figure 4 I Predicting Changes in Gold Medals**

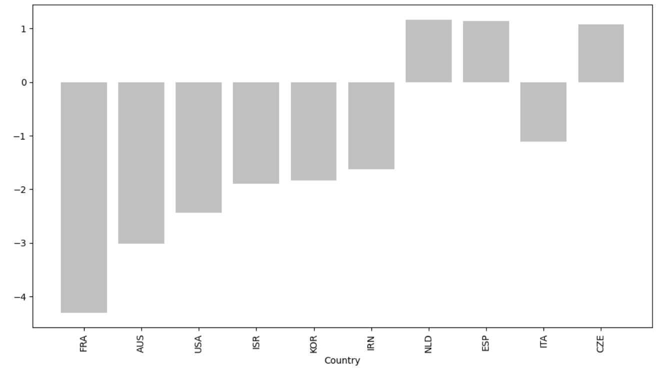


**Figure 5 I Predicting Changes in Silver Medals**

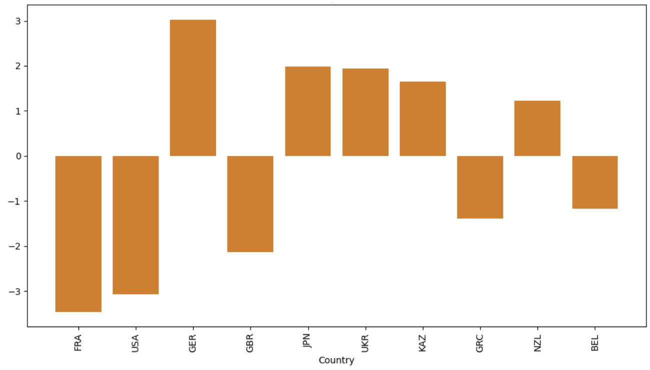


**Figure 6 I Predicting Changes in Bronze Medals**

ical and economic challenges, still has significant potential in traditional sports like gymnastics and athletics. Brazil's continued investment in sports after the 2016 Rio Olympics, especially in football and volleyball, suggests improved performance in 2028.

However, some countries may face challenges and see a decline in performance. Japan, despite its success in the 2020 Tokyo Olympics, may struggle in gold medal counts due to limitations in certain events, particularly in swimming and athletics where other countries are catching up. The UK, while steadily improving since the 2012 London Olympics, may experience pressure on its gold medal count due to the retirement of older athletes and the rise of new competitors. Australia, traditionally strong in swimming and other events, may face increased competition, potentially leading to a decrease in total medal count.

Firstly, we use regression models to predict the number of gold, silver, and bronze medals that each country may receive in 2028.Defining the year as Y and the country code as N, trained regression models, The number of medals won by a country in 2028 are f(x),which can be expressed using the following formula:

$$G_{2028} = f_{Gold}(Y, N) \tag{5}$$

$$S_{2028} = f_{Silver}(Y, N) \tag{6}$$

$$B_{2028} = f_{Bronze}(Y, N) \tag{7}$$

Secondly, for the medal prediction results of each country, we use Monte Carlo simulation to generate the possible distribution of medal numbers.Assuming that the number of gold, silver, and bronze medals follows a normal distribution, simulate whether each country has won its first medal. The formula can be described as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{9}$$

If a country has zero medals in all categories, it is considered to have won its first medal in 2028.The specific calculation formula is:

$$P(\text{first medal}) = \int_R P_{Gold} \times P_{Silver} \times P_{Bronze} \tag{10}$$

Finally, based on the combination of Monte Carlo simulations and the random forest regression model, we predict that two countries will win their first medals at the 2028 Los Angeles Summer Olympics. The standard deviation of this result is 0.1075, indicating that while there is some uncertainty in the prediction, the values remain stable in most cases.

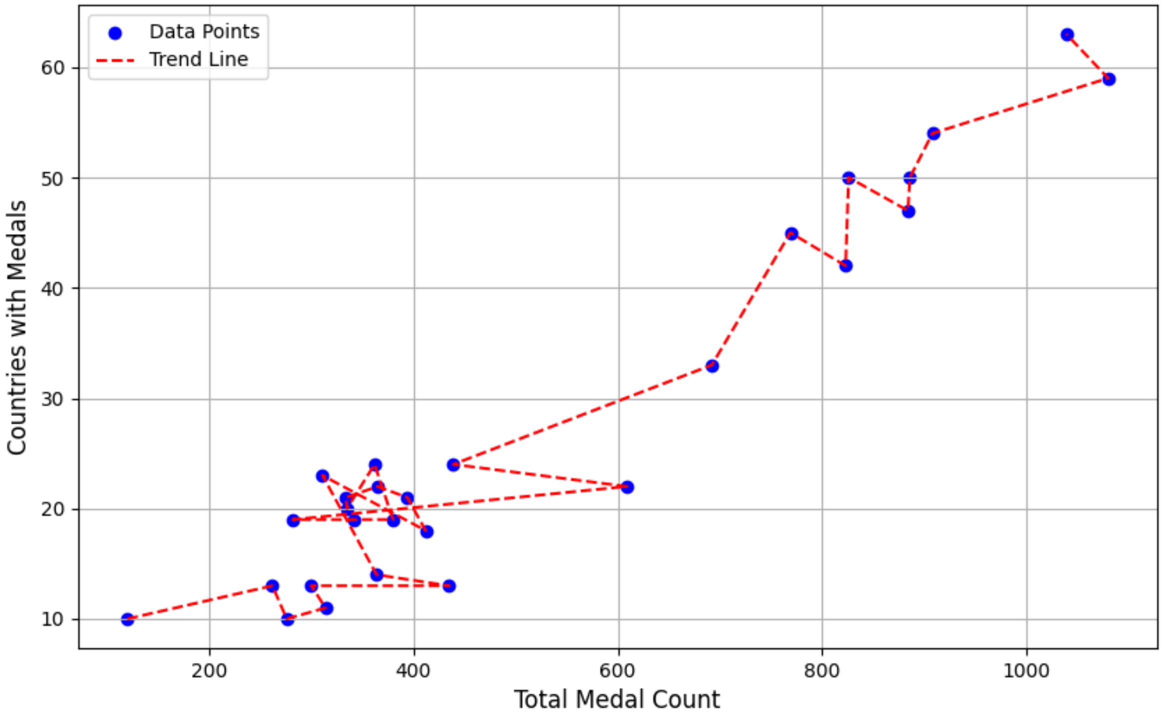


**Figure 7 l Medal-winning Countries vs. Total Medals**

Additionally, we analyzed the relationship between the total number of Olympic medals and the number of medal-winning countries. As shown in the following figures, we observe that as the total medal count increases, the number of countries winning medals also rises. As more events are added, more countries have the opportunity to participate and win medals and they also gaining a chance to excel in certain specific events.

## QUANTITATIVE IMPACT OF THE "GREAT COACH" EFFECT ON OLYMPIC PERFORMANCE

In this section, we explore the relationship between the number and types of Olympic events and the number of medals won by each country. The core issue is to analyze which projects are most important for each country, and how the host country's selection of projects affects medal results[2]. The construction of this model is based on the observation that different movements have varying degrees of importance in different countries, typically depending on the country's historical performance, cultural background, and allocation of training resources. In addition, the host country's selection of new projects is particularly prominent, as they may be related to the host country's own advantageous projects or resource allocation.

We start with the following model to calculate the weighted medal scores for each country in each event:

$$W_{s,c} = \frac{M_{s,c}}{T_c} \tag{11}$$

This weighted score is used to identify the importance of a particular event in the overall medal tally of the country. Countries that achieve a higher medal ratio in a specific event will have a higher weighted score.

Next, we calculate the standardized medal scores for each event. The normalized score is calculated from the ratio between the weighted score of the project and the total number of medals won by all countries in the event:

$$S_{s,} = \frac{M_s}{T_s} \tag{12}$$

For the host country effect, we extracted projects that were modified by the host country from 2000 to 2024 and labeled them as host countries. We used the ARIMA model to fit the data of the number of medals changing over time and matched it with the corresponding year's real data. The core idea of the ARIMA model is to predict time series using a combination model consisting of autoregression (AR) and moving average (MA). The mathematical expression of ARIMA model is:

$$(1-\phi_1 B)(1-B)X_t = \theta_1 B_t + \varepsilon_t \tag{12}$$

The results of medal data from various countries indicate that the top five sports include football, volleyball, triathlon, shooting, and water polo.

The dominant position of these sports, especially in countries such as Brazil, the United States, and China,

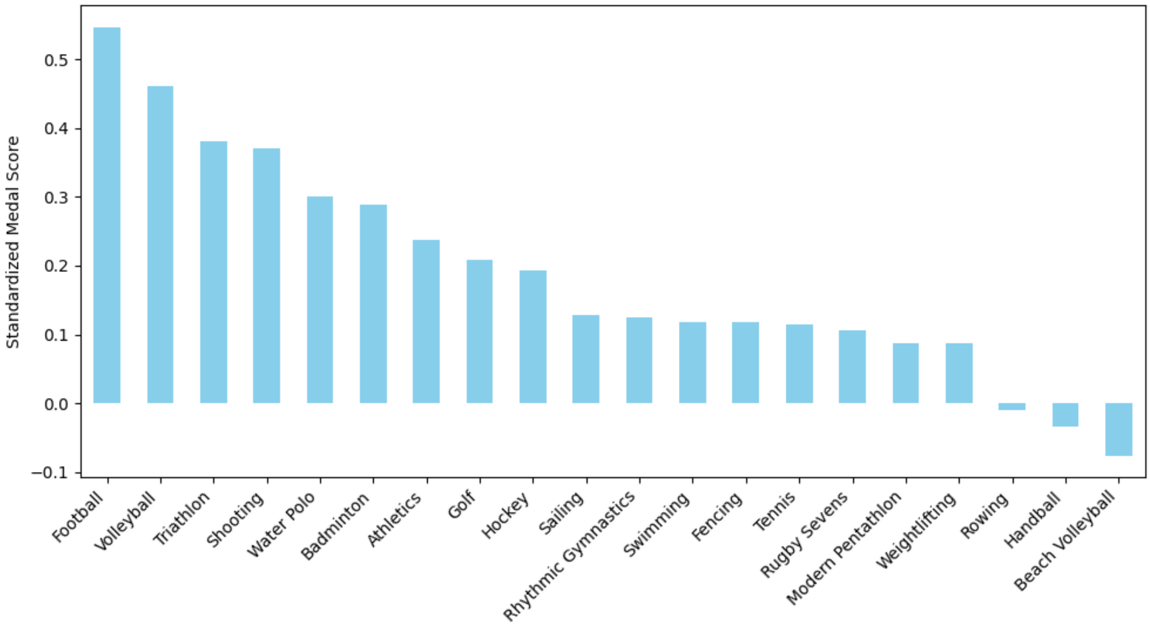


**Figure 8 I Top Sports by Weighted Medal Score**

**Table 1 I SARIMAX Model Regression Data**

| Parameter | Coefficient | Std Err | z-value | P-value | Confidence Interval (95%) |
|---|---|---|---|---|---|
| ar.L1 | -0.4340 | 2.647 | -0.164 | 0.870 | [-5.621, 4.753] |
| ma.L1 | 0.9989 | 65.588 | 0.002 | 0.999 | [-12.845, 12.865] |
| sigma2 | 9.6711 | 632.666 | 0.002 | 0.999 | [-12.400, 12.400] |

can be attributed to their historical investment and emphasis on these events.

In addition, the selection of sports events has a significant impact on the distribution of Olympic medals, especially under the influence of the host country. Host countries usually choose sports projects based on their historical advantages, with the aim of improving their performance in specific fields. For example, Brazil's success in volleyball is reflected in their consistent selection of this sport for high-level competition. Similarly, the United States has significant advantages in athletics and basketball, while China excels in swimming and diving, reflecting the direct impact of the host country's sports choices on the results **(Figure 8)**.

Through model estimation, the host effect has increased the number of medals by an average of about 16% annually since 2000, indicating that this effect has a positive impact on the number of medals.

Ar.L1 (autoregressive coefficient): -0.4340, indicating a negative relationship with the previous moment, but the p-value is 0.870, indicating that the coefficient is not statistically significant.Ma.L1 (moving average coefficient): 0.9989, indicating that the current number of gold medals is greatly affected by previous errors, p-value is 0.999, indicating that the coefficient is not significant.Sigma 2 (residual variance): 9.6711, p-value is 0.999, indicating that the contribution of residual variance to model fitting is weak.The Ljung Box test result shows p-value=0.78, indicating that there is no significant autocorrelation in the residuals, which is consistent with the white noise hypothesis; The Jarque Bela test result has a p-value of 0.90, indicating that the residuals follow a normal distribution.

Although the residual performance of the model is good, the significance of the coefficients is poor, indicating that the model may have been overly simplified and failed to effectively capture the complexity of the data.

## Divergence in Medal Distribution and Evolution of Athlete Performance

The influence of a great coach extends beyond athlete training and competition strategies; it also involves team management, psychological development, tactical planning, and other aspects. Many national teams worldwide have experienced the "Great Coach" effect, where exceptional coaches have helped their teams achieve results that exceeded expectations. To investigate the contribution of this effect to medal counts, we have chosen coaches with significant impacts in Olympic history, along with their corresponding nations.

To calculate the impact of the "Great Coach" effect on medal counts, we have employed a linear regression model. Regression models are commonly used statistical methods for analyzing the relationship between an independent variable and a dependent variable. Our model is as follows:

$$Y_i = \beta_0 + \beta_1 \cdot \textit{Coach Period}_i + \varepsilon_i \quad (13)$$

in this model, the key parameter of interest is $\beta_1$, which represents the impact of the “Great Coach” effect on the medal count. If $\beta_1 > 0$, it suggests a positive influence of the “Great Coach” on medal counts; conversely, if $\beta_1 < 0$, it implies a minimal or negative effect.

To obtain the data required for our model, we selected Brazil, the United States, and China as the focus countries. We filtered the Olympic data for these countries to extract the annual medal counts for gold, silver, and bronze, and tagged each year to indicate whether it falls within the “Great Coach” period.

Specifically, the "great coach" effect of American women's gymnastics occurred from 2008 to 2016, while the "great coach" period of the Brazilian women's volleyball team was from 1994 to 2000. Lang Ping coached the American women's volleyball team from 2008 to 2016 and the Chinese women's volleyball team from 2016 to 2021. These time periods allow us to add a “Great Coach” label for the corresponding year in each dataset.

To begin, we aggregated the medal data for each country, focusing on the periods when the selected coaches were in charge. Specifically, we extracted data from Brazil’s women’s volleyball, the U.S. women’s volleyball, and China’s women’s volleyball teams, which had clear periods of coaching by well-known figures like Lang Ping and Bernardo Rezende[3]. The following steps were executed:

1) We collected the medal counts (gold, silver, and bronze) for each country’s team on a per-year basis, while incorporating the years that fall under the "Great Coach" period into the dataset.
2) We applied the regression model to estimate the impact of the "Great Coach" effect. Specifically, we analyzed how the medal counts changed under the guidance of the coaches during their respective periods. The regression results provide us with an estimate of the effect’s significance on gold, silver, and bronze medals.
3) The analysis of the regression outputs helped us quantify how much the coaching periods influenced medal counts, the $\beta_1$ parameter of the regression model indicated the extent of this influence during the "Great Coach" periods .

The regression outputs yielded significant insights into how the "Great Coach" effect has shaped Olympic medal outcomes. For each country and sport analyzed, we found that the "Great Coach" periods were associated with positive shifts in medal counts, particularly in volleyball, where Brazil's and the U.S.'s performances were significantly improved during the tenure of their respective great coaches.

To aid in the visualization of these results, we also plotted the historical trends of medal counts for each country and highlighted the periods in which "Great Coaches" were active. The graphs below provide a clear representation of how medal counts changed before, during, and after the coaching periods of these influential figures.

**Figure 9** show the impact of the "Great Coach" effect on the corresponding project, The light green background represents the "Great Coach" effect period, with dots and crosses of three colors representing actual and predicted medal counts.

These images show a comparison between the actual medal count and the predicted medal count, combined with the 'coaching effect' at different time periods. Each chart represents a specific country or sport, displaying predicted and actual medal results for different years[4]. The highlighted green area represents the period when the "Great Coach" effect is active, emphasizing the potential impact of coaches on medal counts. We observed that after introducing the 'great coach', the number of medals increased significantly, and the actual number of medals exceeded the predicted value. The US women's volleyball team presents a broader historical perspective, with medal numbers initially low but significantly increasing under the influence of the 'great coach', and actual medal numbers exceeding predictions. Finally, the Chinese women's volleyball team provided a comprehensive comparison, highlighting the overall impact of coaches on medal counts, with the green shadow period coinciding with the existence of the "great coach" effect. These charts collectively demonstrate how the influence of 'great coaches' significantly changes medal predictions, during which actual results often exceed expectations.

**Table 2** indicates that the great coach effect does have an impact on the medal count of these events, with the impact on team events limited to only one team having a smaller impact, while individual events such as gymnastics have a more significant impact. But most of the regression results did not reach a significant level. Therefore, although 'great coaches' may bring some positive changes, they lack statistical significance from the perspective of regression equations.

In order to address the issue of selecting countries and determining which sports should consider hiring 'great coaches', we first screen historical data for medal changes and select countries and events with significant fluctuations during specific time periods. Firstly, we choose competition data between 1992 and 2024, especially for team events. Considering that the selected countries are relatively weak in terms of events, medal analysis will not be conducted. Instead, by analyzing previous ranking data, some countries and events with significant fluctuations will be selected. Then, we used

**Table2 I Regression Data of the "Great Coach" Effect**

| Sport Event | Coef (Const) | Coef(Period) | R-squared | Adj. R-squared | P>\|t\| |
|---|---|---|---|---|---|
| BRA-Women Volleyball | 0.2 | -0.2 | 0.040 | -0.056 | 0.533 |
| BRA-Men Volleyball | 0.0833 | 0.4167 | 0.214 | 0.158 | 0.0714 |
| USA-Women Volleyball | 0.1 | -0.1 | 0.025 | -0.064 | 0.606 |
| CHN-Women Volleyball | 0.2222 | 0.2778 | 0.058 | -0.047 | 0.476 |
| USA- Gymnastics | 2.1579 | 2.5921 | 0.054 | 0.009 | 0.286 |

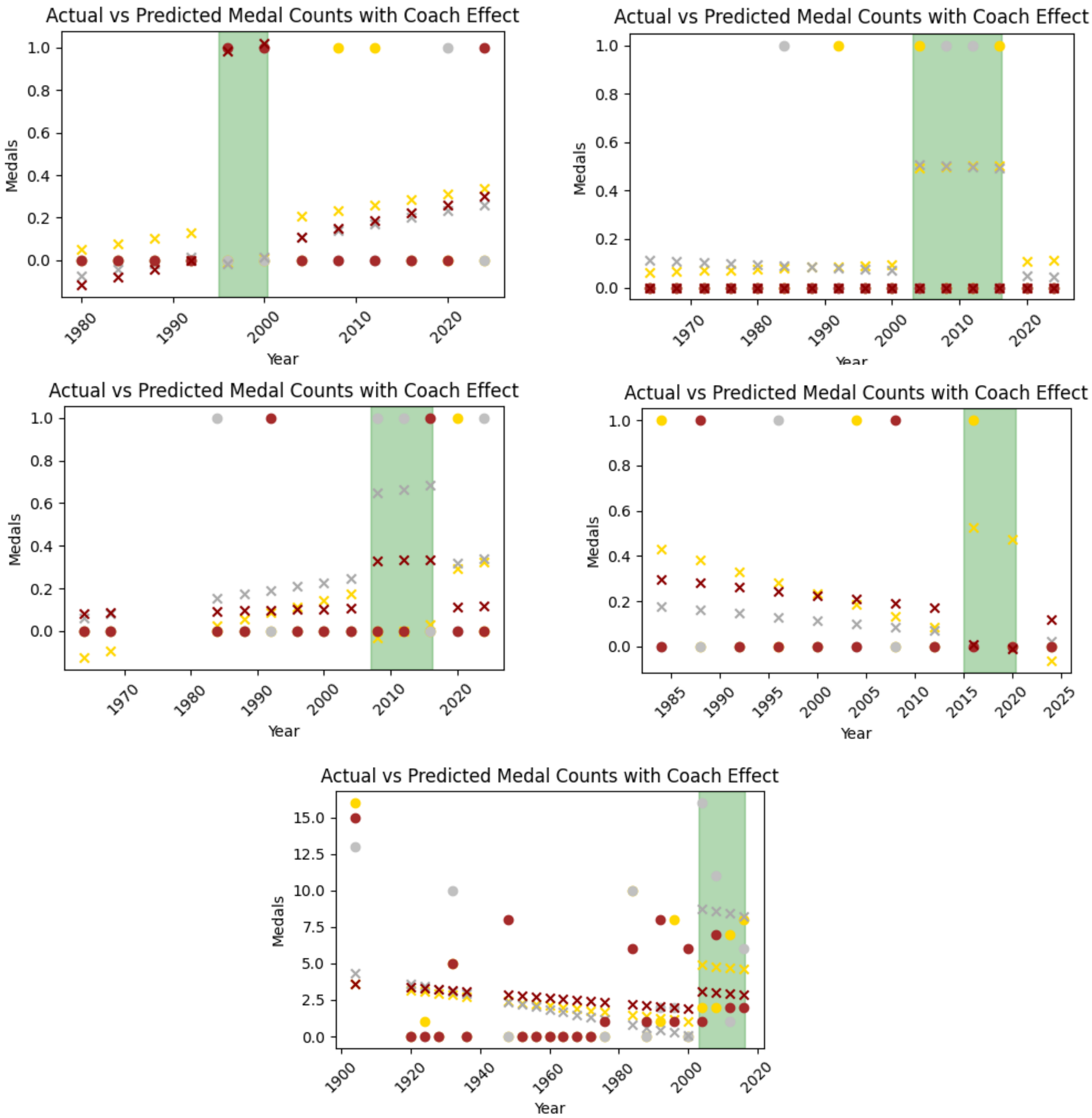


**Figure 9 I "Great Coach" Effect Contrast**

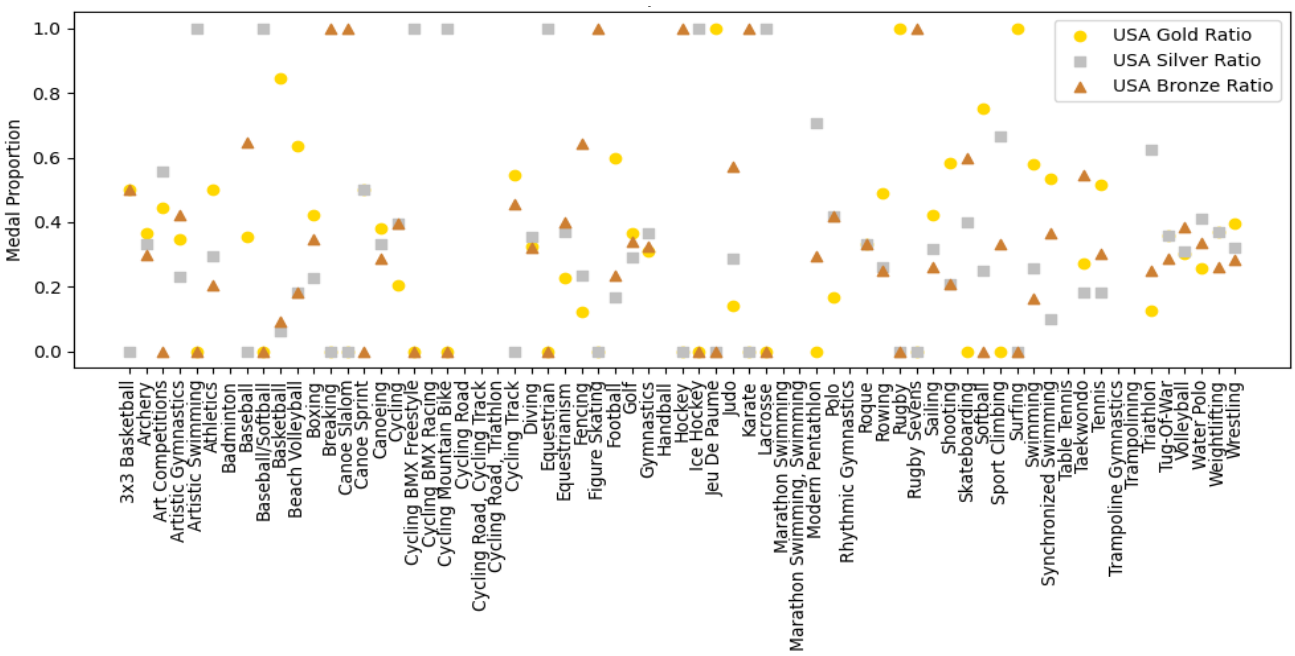


**Figure 10 I Distribution of Medals in the USA**

the OLS regression model to hypothesize the 'Great Coach Effect', selecting the top three countries most affected by it and corresponding projects: football in the United States, basketball and football in China, and basketball in Russia. Finally, for the United States, the relevant p-value is 0.1705, for China it is 0.2778, and for Russia it is 0.4167. But the rankings of the corresponding projects in the three countries have risen by 5-21 places.

This indicates that although the impact of the "great coach" effect is not certain from a statistical perspective, we can consider that hiring a "great coach" can indeed to some extent compensate for the weaknesses of these countries in these projects.

## Other Findings

It can be seen from the chart that there are significant differences in the medal ratios of the United States, the United Kingdom, and China in various sports.

The United States excels particularly in sports such as swimming, athletics, and basketball, primarily due to its long-standing sports tradition, robust training system, and abundant resource support. The United States boasts world-class swimming and athletics training facilities, nurturing a large number of elite athletes, with its swimming team consistently achieving remarkable results in the Olympics and World Championships. Furthermore, the NBA league in American basketball is not only the top basketball league globally but also supplies a significant number of outstanding talents to the national team, enhancing the dominance of American basketball on the international stage **(Figure 10)**.

In contrast, the United Kingdom performs exceptionally well in cycling, rowing, and athletics. The emphasis and investment in cycling in the UK are particularly notable, with comprehensive training facilities and top-notch coaching teams, enabling the British cycling team to win multiple gold medals in the Olympics and World Championships. The rowing events in the UK are equally impressive, closely linked to its rich rowing history and deep cultural heritage, with many renowned rowing clubs and competitions providing excellent training and competition environments for athletes **(Figure 11)**.

China dominates in diving, gymnastics, table tennis, and weightlifting. The Chinese diving team is renowned for its superb technique and consistent performance, attributed to its rigorous training regime and scientific training methods. In gymnastics, the Chinese team has achieved outstanding results in international competitions, demonstrating strong competitiveness. In table tennis, China's advantage is indisputable, closely related to the extensive grassroots support and excellent training system within the country, with many outstanding table tennis players beginning professional training at a young age, contributing a significant number of talents to the national team **(Figure 12)**.

By analyzing Phelps' medal achievements in conjunction with the images, it can be observed that his performance from 2000 to 2016 was exceptionally outstanding **(Figure 13)**. Firstly, as seen in the first chart, the number of gold medals Phelps won steadily increased from 2000 to 2008, peaking in 2008 with a total of 8 gold medals. This period marked the pinnacle of his career, particularly during the 2008 Beijing Olympics, where he remarkably secured 8 gold medals and broke multiple world records. However, starting in 2010, there

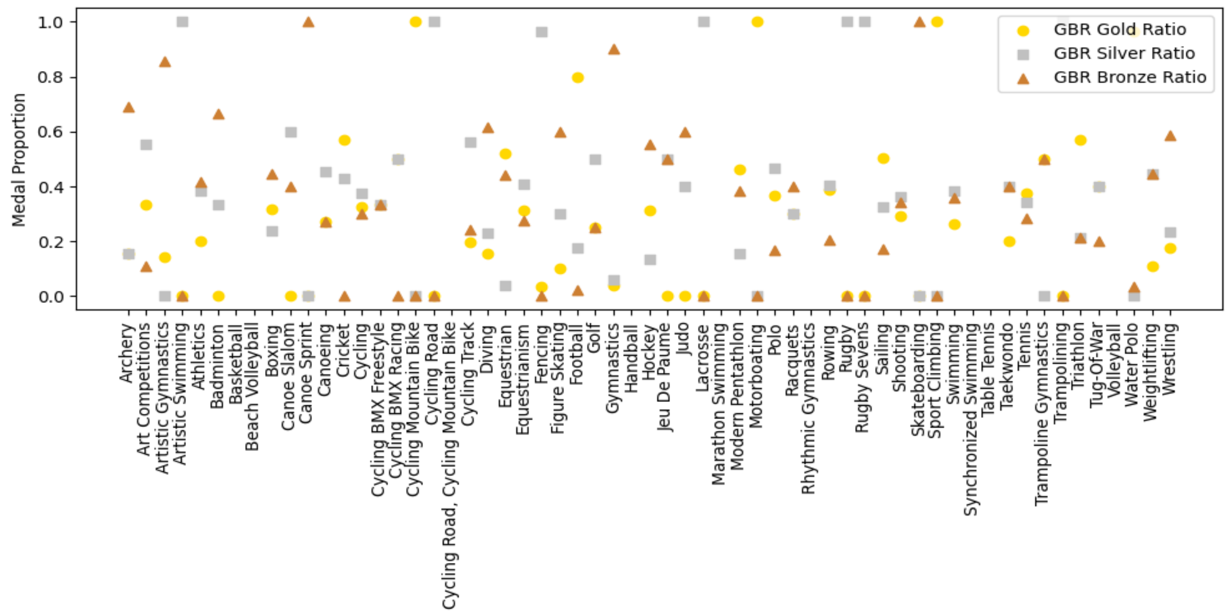


**Figure 11 | Distribution of medals in the GBR**

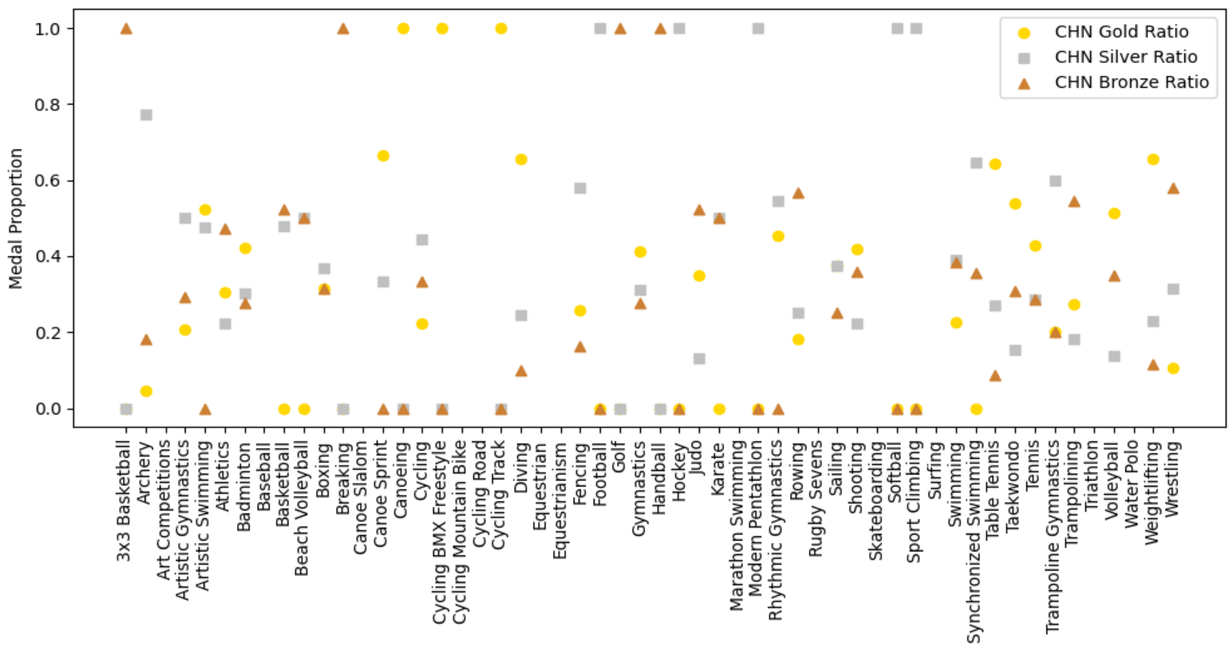


**Figure 12 | Distribution of Medals in the CHN**

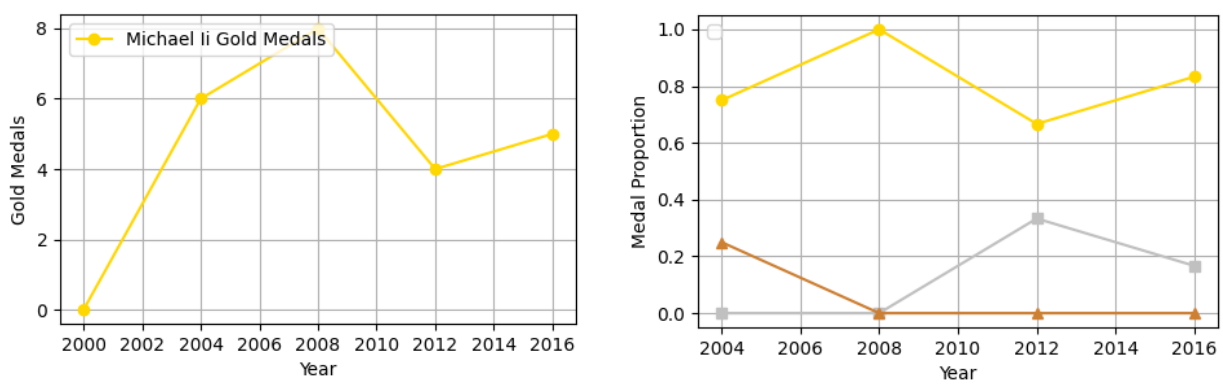


**Figure 13 | Phelps' Medals Change Over Time**

was a decline in his gold medal count, which may be attributed to aging, changes in physical condition, and intensified competition. Nevertheless, he maintained a high level of performance in 2012 and 2016, winning 4 and 5 gold medals respectively. This indicates that Phelps not only possesses exceptional athletic ability but also demonstrates strong adaptability and enduring competitiveness.

The second chart illustrates the ratio of gold, silver, and bronze medals Phelps earned in different years. It is evident that his gold medal ratio reached its peak in 2008, with nearly all of his medals being gold. This further confirms his dominance during that year. In 2012 and 2016, although the ratio of gold medals declined, he still secured a considerable number of silver and bronze medals, showcasing his versatility and consistency across various events.

Phelps' success is attributed to his rigorous training regimen, a scientific management team, and strong psychological resilience[5]. His medal achievements reflect his competitive state and adaptability at different stages. His dominance during peak periods, perseverance during low points, and stable performance in the later stages of his career all exemplify his extraordinary qualities and remarkable accomplishments as a top athlete. His outstanding performance in swimming not only earned him numerous honors but also set a benchmark for future swimmers.

## STRENGTHS AND WEAKNESSES

### Strengths

- This model mainly uses traditional mathematical modeling methods such as random forest regression, time series analysis, and Monte Carlo simulation to predict Olympic medal results. Combine the advantages of each algorithm during the processing.
- Based on objective reality and official data files, an analysis was conducted on the award situation of most athletes and countries in the 33rd Olympic Games. Evaluated, optimized, and enhanced the authenticity of the model.
- Based on practical problems and needs, analyzed the advantages and disadvantages of specific countries in their sports projects.
- Using stage precision isolation to shorten the running time of the algorithm when the values can reflect reality as much as possible.

### Weaknesses

- The limitations of certain assumptions may limit the consideration of practical situations, especially for research dealing with dynamic and politically sensitive areas such as the Olympics.
- The impact of host country event selection was mentioned, but there is a lack of larger data validation.
- Limited exploration of external factors: This analysis may benefit from a broader discussion of external factors.
- Although Monte Carlo simulations are used to quantify uncertainty, some results still exhibit significant differences.

**References**

1. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
2. Chatterjee, S., et al. (2012). Statistical Analysis of Olympic Medal Counts: Predicting Performance. Journal of Sports Analytics.
3. Rynne, S. B., Mallett, C., & Tinning, R. (2006). High Performance Sport Coaching: Institutes of Sport as Sites for Learning. *International Journal of Sports Science & Coaching*, *1*(3), 223-234.
4. Chatterjee, S., et al. (2012). Statistical Analysis of Olympic Medal Counts: Predicting Performance. Journal of Sports Analytics.
5. Song, Y., & Lan, H. (2024). The effects of high-intensity interval training on cardiometabolic health in children and adolescents: A systematic review and meta-analysis. *Journal of Sports Science and Medicine, 23*(4), 690-706.