Problem Chosen	2025 MCM/ICM Summary Sheet	Team Control Number
C		2511431

DGBIF-OMF: Olympic Medals & Strategy

Summary

The purpose of this paper is to establish a prediction model to analyze and forecast the medal distribution for the 2028 Los Angeles Summer Olympics. The model specifically focuses on three main issues: the prediction of medal distribution among countries, the impact of the "great coach" effect on medal distribution, and the influence of event settings on medal distribution.

Firstly, we constructed a Dynamic Grey-Bayesian Integration Framework (DGBIF-OMF) model to predict the medal distribution for the 2028 Olympics. This model takes into account multiple factors such as athlete performance, national historical trends, host advantage, and event sensitivity. By analyzing historical data, we forecasted the total number of medals and the counts of each type of medal for various countries, along with prediction intervals.

Moreover, the model assessed the uncertainty and accuracy of the predictions, as well as model performance. Utilizing this model, we determined that the top three countries in terms of total medals at the 2028 Los Angeles Olympics would be the United States, China, and Australia. Among them, Antigua and Barbuda, Guinea-Bissau, and three other countries are highly likely to win their first Olympic medals at the 2028 Los Angeles Olympics.

Secondly, we explored the potential impact of the "great coach" effect on medal distribution. By analyzing the Olympic careers of two outstanding coaches, we introduced an excellent coach effect coefficient and used the sliding window method and Monte Carlo simulation to estimate its impact on the number of medals. Based on this, we selected China, Japan, and the United Kingdom as three countries to recommend projects worth investing in hiring coaches and quantified the impact of coaches' cross-country coaching on the medal contributions of these countries.

Finally, based on our analyses, we recommend that NOCs consider event setting changes for better resource allocation, strategy formulation, and enhanced medal competitiveness. Countries should focus on events matching their strengths and potential new or adjusted events in the next Olympics. This approach can optimize resource use and boost Olympic medal competitiveness.

The research findings of this paper not only provide a scientific basis for National Olympic Committees to optimize their preparation strategies, but also lay the foundation for studying the changes in the international sports landscape and its potential underlying patterns.

Keywords: Olympic Medal Forecasting; Dynamic Grey-Bayesian Model;"Great Coach" Effect

Contents

1 Introduction

1.1 Problem Background

As one of the most prestigious global sporting events, the Olympic Games have garnered significant attention regarding medal distribution and performance trends. The Games not only highlight the dominance of traditional sports powerhouses such as the United States and China but also underscore their universality and global impact, as evidenced by smaller nations like Albania and Dominica winning their first medals at the 2024 Paris Olympics.

Historical data reveals substantial temporal variations in medal counts and rankings across participating nations. These variations are driven by a multitude of factors, including the emergence of star athletes, the home advantage effect, the introduction of new events, and the strategic influence of elite coaching. By systematically analyzing and predicting Olympic medal distributions through mathematical modeling, we can uncover underlying patterns in medal trends and provide actionable insights for national Olympic committees to optimize their preparation strategies.

1.2 Restatement

Based on the provided background information and dataset, we address the following key questions:

Problem 1: Medal Distribution Prediction and Trend Analysis Using data on the number of athletes and events participated in by each country, develop a predictive model to estimate the medal distribution for the 2028 Summer Olympics in Los Angeles. Analyze historical trends in medal performance across countries and quantify the probability of a first-time medal win for nations that have not previously won an Olympic medal.

Problem 2: Impact of Coaching on Medal Distribution Investigate the influence of high-quality coaching on Olympic medal outcomes. Identify three countries that would benefit most from investing in elite coaching programs and recommend specific projects for implementation. Additionally, quantify the impact of cross-border coaching on a nation's medal contribution.

Problem 3: Medal Distribution Patterns and Strategic Recommendations Synthesize the modeling approaches and results from Problems 1 and 2 to identify underlying patterns in Olympic medal distribution. Provide actionable theoretical guidance for National Olympic Committees (NOCs) to optimize resource allocation, develop competitive strategies, and enhance medal-winning potential.

1.3 Literature review

Since the post-World War II era, sociologists and economists have extensively studied the relationship between socioeconomic factors and Olympic medal outcomes. Johnson and Ali (2004) proposed a linear model to predict medal counts based on GDP per capita, population size, and binary variables representing host country status and political system. Building on this framework, Nader and Booth (2008) introduced a Cobb-Douglas production function to account for the fact that many countries do not win medals in a given Olympic Games. Their model incorporated population, economic resources (measured by GDP), and a host country advantage dummy variable. By including a lagged dependent variable to capture a nation's historical medal performance, they achieved a "reasonably accurate" prediction of medal tallies for the Sydney 2000 Olympics.

More recently, Cha and Pershing (2015) analyzed the determinants of medal counts by examining the distribution of disciplines (number and type) in which countries excel. Their findings revealed that higher-income nations tend to exhibit less specialization, winning medals across a broader range of sports.

This study establishes the DGBIF-OMF framework by integrating grey prediction with Bayesian modeling, which synthesizes athlete performance, historical medal rankings, host nation advantages, and event quantity/type variations to forecast the 2028 Olympic medal table.

1.4 Our Work

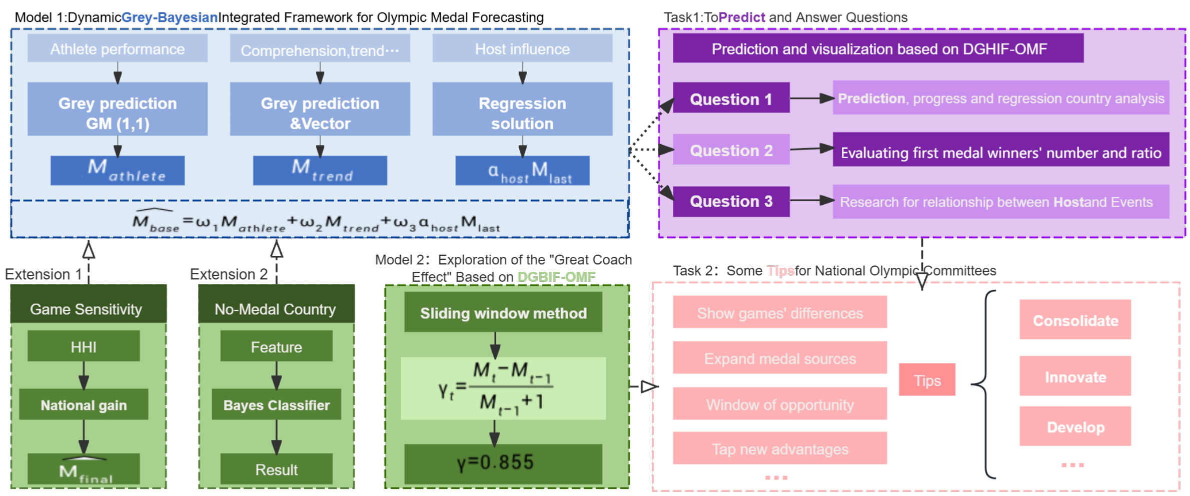

Figure 1: Overview of our work

2 Assumptions and Justifications

▼ **Hypothesis 1: Continuity in Medal Trends.** Medal trends exhibit continuity over time, with historical performance serving as a reliable predictor of future outcomes.

▲ **Rationale:** Stable sports development systems and long-term dominance in specific disciplines support consistent medal trends.

▼ **Hypothesis 2:Athlete Participation Limit.** Athletes participate in a maximum of four Olympic Games during their careers.

▲ **Rationale:** Most athletes reach their peak performance within a 12–16 year window, aligning with four Olympic cycles.

▼ **Hypothesis 3:Stable Host Country Advantage.** The host country maintains a stable medal advantage due to home support and venue familiarity.

▲ **Rationale:** Historical data shows a consistent increase in medal counts for host nations.

▼ **Hypothesis 4: Stable Athlete Participation.** Athlete participation in specific sports remains stable, with historical performance predicting future results.

▲ **Rationale:** Athletes typically specialize in specific disciplines, leading to consistent performance patterns.

3 Notations

Table 1: Notations used in this paper

Symbol	Description	Unit
M_{ij}	Number of gold, silver, bronze, and total medals won by country i in the j-th Olympic Games (vector)	
M_{host}	Host country impact factor	
HHI_i	Herfindahl-Hirschman Index to measure the concentration of medals for country i	
ω_i	Weight of the i-th factor in the prediction model	
γ	Excellent coach benefit coefficient	
$\lambda(k)$	Ratio of the $(k-1)$-th data point to the k-th data point	
W_{ij}	Historical event concentration, the proportion of medals in event j to all medals of country i	

4 Model 1:Dynamic Grey-Bayesian Integrated Framework for Olympic Medal Forecasting (DGBIF-OMF)

4.1 Data Description

Handling of Data Continuity: Geopolitical changes in history, like the Soviet Union's dissolution, disrupted the continuity of sports - related data. To address this, we performed a data - mapping process. For Russia, historical records from the Soviet era were mapped to its current entries, ensuring data consistency across different periods. This provided a reliable basis for subsequent analysis. To maintain research rigor, we also applied this data succession method to the following countries(Table 2):

Table 2: Country Name Conversion Table

NOC - Pre	NOC - Changed
FRG	GER
GDR	GER
YUG	SRB/MNE
URS	RUS
TCH	CZE
YAR	YEM
...	...

Data Cleansing and Standardization: To address potential issues in merging tables caused by inconsistent country naming across datasets, we standardized the country names. For instance, "United Kingdom" and "Great Britain" were unified as "United Kingdom." This standardization effectively mitigates risks of double counting and minimizes confusion arising from naming discrepancies, thereby enhancing the accuracy and coherence of the data.

Creation of Unique Identifiers: To analyze and compare performance data across years, we used athlete name, event, and year as primary keys to uniquely identify each record. To visualize changes over multiple Olympic cycles, data were grouped by athlete name and event, then sorted by year in descending order, prioritizing the most recent performances and highlighting trends.

4.2 The Establishment of Model 1

DGBIF-OMF: Core Architecture Construction

As numerous factors influence the Olympic medal table, we aim to analyze their impacts step by step, progressing from the "backbone" to the "branches." First, a basic modeling framework incorporating athlete performance, national historical trends, and host nation effects will be established.

$$\widehat{M_{\text{base}}} = \omega_1 M_{\text{athlete}} + \omega_2 M_{\text{trend}} + \omega_3 a_{\text{host}} M_{\text{last}} \tag{1}$$

To construct this base model, we will develop predictive sub-models for athlete performance, national medal trends, and host influence factors in the following sections.

Performance-Driven Athlete Prediction Sub-Model

Athletes are key to medal - winning. We propose assessing their medal prospects in the next Olympics based on past event performances, then estimating a country's total medal count. Thus, we'll build a gray prediction model using athletes' past medal records.

We quantified the awards for each event in which athletes participated during previous Olympic Games and recorded them as columns of data. Subsequently, we applied the gray prediction algorithm to model the data(As shown in the Figure 2).

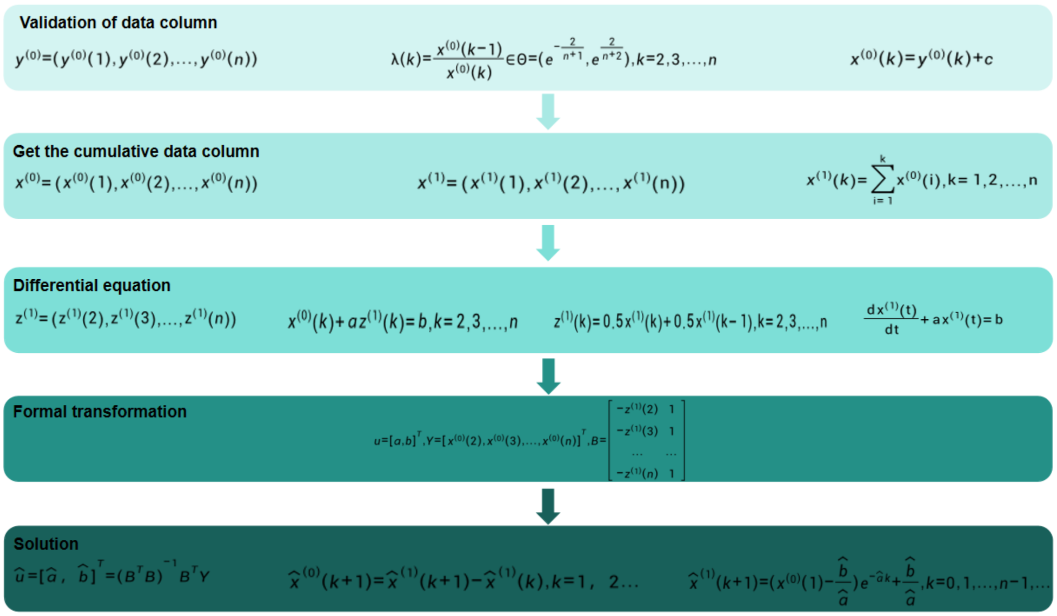

Figure 2: Flowchart of the Grey Prediction Algorithm

By applying the above method, we calculated the awards obtained by each athlete in various events during the jth Olympics and derived prediction results M_{athlete} based on their performance.

Historical Trend-Driven National Prediction Sub-Model

While the performance of individual athletes is crucial, the overall strength of a country cannot be overlooked. A country's overall strength often determines the level of resources it invests in sports, which significantly increases its likelihood of winning medals at the Olympics. By analyzing the historical trends of a country's medal achievements, we can indirectly evaluate its overall strength and predict its medal performance in future Olympic Games. Based on this analysis, we constructed a grey prediction model to assess the relationship between a country's historical performance and its future awards. Specifically, we represent the number of gold, silver, and bronze medals, as well as the total medal count for each country in previous Olympics, as data vector columns.

$$S^{(0)} = \left(S^{(0)}(1), S^{(0)}(2), \dots, S^{(0)}(n)\right) \tag{2}$$

$$S^{(0)}(1) = \left(s^{(0)}_{\text{Gold}}(1), s^{(0)}_{\text{Silver}}(1), s^{(0)}_{\text{Bronze}}(1), s^{(0)}_{\text{Total}}(1)\right)^T \tag{3}$$

$$s^{(0)}_{\text{Gold}}(1) + s^{(0)}_{\text{Silver}}(1) + s^{(0)}_{\text{Bronze}}(1) = s^{(0)}_{\text{Total}}(1) \tag{4}$$

The remaining steps follow the same prediction process as that based on athletes' performance. By applying this method, we can predict each country's medal count for the jth Olympic Games:

$$M_{\text{trend}} = \left(s_{Gold}^{(0)}(j), s_{\text{Silver}}^{(0)}(j), s_{\text{Bronze}}^{(0)}(j), s_{\text{Total}}^{(0)}(j)\right) \tag{5}$$

Host Effect-Driven Olympic Prediction Sub-Model

In addition to the significant influence of athlete performance and national historical trends on Olympic medal outcomes, we posit that the host country factor plays a pivotal role in determining medal distribution. Hosting the Olympic Games confers a distinct home field advantage to the nation's athletes, eliminating potential performance impediments such as jet lag and climate adaptation. To quantify this phenomenon, we construct a linear regression model that examines the relationship between the medal count of the host nation during its hosting session (denoted as session j) and its performance in the two adjacent Olympic sessions. Our model is predicated on the assumption of a linear relationship between these variables, expressed as follows:

$$s^{(0)}(j) = 0.5a_{\text{host}}\left(s^{(0)}(j-1) + s^{(0)}(j+1)\right) \tag{6}$$

$$S^{(0)}(j) = \left(s_{\text{Gold}}^{(0)}(j), s_{\text{Silver}}^{(0)}(j), s_{\text{Bronze}}^{(0)}(j), s_{\text{Total}}^{(0)}(j)\right)^{T} \tag{7}$$

$$a_{\text{host}} = \left(a_{\text{host}}(\text{ Gold }), a_{\text{host}}(\text{ Silver }), a_{\text{host}}(\text{ Bronze }), a_{\text{host}}(\text{ Total })\right) \tag{8}$$

We can solve for the host impact factor for all host countries:

$$a_{\text{host }(j)} = \frac{2S^{(0)}(j)}{S^{(0)}(j-1) + S^{(0)}(j+1)} \tag{9}$$

These are then averaged to get the final host impact factor:

$$a_{\text{host}} = \frac{1}{k}\sum_{j=1}^{k} a_{\text{host }(j)} \tag{10}$$

According to the above process, sovle for:

$$c \geq 9.5 \tag{11}$$

$$a_{\text{Gold}} = 1.9154, a_{\text{Silver}} = 1.6685, a_{\text{Bronze}} = 1.3993, a_{\text{Total}} = 1.7257 \tag{12}$$

When the country is the host of the predicted session, the parameters are:

$$\widehat{M_{\text{base}}} = 0.44M_{\text{athlete}} + 0.38M_{\text{trend}} + 0.18\text{a}_{\text{host}}\, M_{\text{last}} \tag{13}$$

When the country is not hosting the projected session, the parameters are:

$$\widehat{M_{\text{base}}} = 0.53M_{\text{athlete}} + 0.47M_{\text{trend}} \tag{14}$$

This yields an integrated, multifaceted base prediction denoted as M_{base}.

4.3 Extension 1:Sensitivity Analysis for Tournaments

Since event factors are an integral part of Olympic medal forecasting, we attempt to explore the relationship between events and medals and analyze which events are most important to different countries. This inquiry not only helps to construct the event sensitivity module, but also further advances the prediction of medal wins for countries when host countries organize new events.

The importance of programs is analyzed first. For each country i and sport j, the historical program concentration is calculated as follows:

$$W_{ij} = \frac{M_{ij}}{M_i} \tag{15}$$

Here, M_{ij}denotes the total number of historical medals won by country i in sport j, and M_i represents the total number of historical medals won by country i across all sports.

$$HHI_i = \sum_{j=1}^{N} W_{ij}^2 \tag{16}$$

A higher value of the HHI indicates a greater concentration of events on which the country relies to win medals (e.g., Kenya's Olympic medals primarily come from long-distance running events). This metric enables us to quantify the relationship between events and medals for each country, thereby analyzing the significance of different events to different countries.

Next, we proceed to construct the event sensitivity module to analyze the impact of host countries introducing new events on the medal counts of various countries and integrate it into the base model. We mainly infer the brand new events that the host country may launch based on historical data, as well as the sub-events in a major category of sports that W_{ij} judge will be launched. Using the probability distribution of these events, we can further estimate the potential increase in medal counts for each country.

The host country's gain is expressed as:

$$\Delta M_{\text{host}} = \sum_{\text{type}} P_{\text{type}} \cdot a_{\text{type}} \cdot \overline{M_{\text{type}}} \tag{17}$$

Among these,P_{type} represents the probability of adding a new project type,$a_{\text{type}} = W_{\text{itype}} \cdot a_{\text{host}}$ is the host's likely performance in this type of program, $\overline{M_{\text{type}}}$ represents the average number of medals for the same type of program.

The non-host gain is expressed as:

$$\Delta M_i = \sum_{\text{type}} P_{\text{type}} \cdot \beta_{\text{type}} \cdot \overline{M_{\text{type}}} \tag{18}$$

Among these,$\beta_{type} = M_{type}$. Based on this, we can expand the tournament sensitiviSSty module within the base model to further refine our comprehensive prediction framework:

$$\widehat{M_{\text{final}}} = \widehat{M_{\text{base}}} + \Delta M_{\text{host}} + \Delta M_i \tag{19}$$

4.4 Extension 2:Predictions for Countries Without Medals

The above model is insufficient to accurately determine the likelihood and probability of a country that has never won a medal achieving its first medal. Therefore, we decided to construct a simple Bayesian model to extend the medal prediction to include non-medal-winning countries.

While the integrated model effectively predicts the medal counts for all countries that have previously won medals, it does not encompass countries that have yet to secure their first medal. To address this, we developed an extension to predict the likelihood of countries winning their first medal, thereby supplementing the integrated model. The results from this extension can be seamlessly incorporated into the original prediction dataset without compromising the model's comprehensiveness or relevance.

For the given data, the countries that won their first medal at each Olympic Games and the countries that participated but have not yet won a medal are represented by the following set:

$$T = \{t_1, t_2, \ldots, t_n\} \tag{20}$$

Next, we define the set of events representing whether a country that has not yet won an Olympic medal achieves its first medal in a given Olympic Games:

$$Q = \{ \text{ Obtain, Not Obtian } \} \tag{21}$$

Next, we define a vector of characteristics for the country whose impact we aim to explore (e.g., number of events participated in, proximity to the host country, number of athletes, etc.). Assuming these characteristics are independently uncorrelated, we have:

$$R = (R_1, R_2, \ldots, R_n) \tag{22}$$

$$P(R=r)=P\left(R_1=r_1\right)P\left(R_2=r_2\right)\ldots P\left(R_n=r_n\right) \tag{23}$$

The formula for the conditional probability is then given by:

$$P\left(R=r \mid Q=c_k\right)=\prod_{j=1}^{n} P\left(R_j=r_j \mid Q=c_k\right) \tag{24}$$

Then, based on Bayes' theorem, we derive a Bayesian classifier:

$$f(r)=\arg\max_{c_k} P\left(Q=c_k \mid R=r\right)=\arg\max_{c_k} \frac{P\left(Q=c_k\right)\prod_j^n P\left(R_j=r_j \mid Q=c_k\right)}{\sum_k^m P\left(Q=c_k\right)\prod_j^n P\left(R_j=r_j \mid Q=c_k\right)} \tag{25}$$

When the feature is r, the conditional probability for all categories is calculated, and the category with the highest conditional probability is selected as the classification result. Since the denominator in the above formula is the same for all categories, it can be disregarded in the calculation, i.e.:

$$f(r)=\arg\max_{c_k} P\left(Q=c_k\right)\prod_j^n P\left(R_j=r_j \mid Q=c_k\right) \tag{26}$$

With this classifier, we can input the country's vector of eigenfactors to calculate both its probability of winning a medal for the first time and its probability of not winning a medal, as follows:

$$P\left(Q=c_k \mid R=r\right)=\frac{P\left(Q=c_k\right)\prod_j^n P\left(R_j=r_j \mid Q=c_k\right)}{\sum_j^m P\left(Q=c_k\right)\prod_j^n P\left(R_j=r_j \mid Q=c_k\right)} \tag{27}$$

This, in turn, provides a prediction of the likelihood and probability that a country yet to win a medal will achieve its first medal.

Some studies [Q] suggest that a country's medal count may be influenced by its proximity to the host country, as athletes from nearby countries may perform better without the interference of jet lag. Building on this consideration, we incorporated a combination of proximity and similarity to the host country's sporting culture into our model. For instance, we found that Antigua and Barbuda (As shown in the Figure 3), one of the countries without any Olympic medals, is geographically close to Los Angeles and shares similarities in sporting culture with the host country. We believe this factor may increase Antigua and Barbuda's likelihood of winning their first medal.

Of course, there is more to consider, and we identified the following factors as key eigenvectors influencing a country's first Olympic medal: number of events, proximity to the host country, and number of athletes.

Figure 3: Geographical location display map

We then quantified and normalized the three characteristic factors, with proximity to the host country quantified by the difference in time zones between the country and the host nation. Subsequently, data from previous Olympic Games were used to calculate the probabilities required for the naïve Bayesian model. The statistically derived probabilities, along with the feature vectors of countries that have not yet won an Olympic medal, were then input into the Bayesian classifier (Equation 26) to determine whether these countries could win their first medal at the next Olympic Games. Their corresponding probabilities were calculated using Equation 27.

5 Task 1:Prediction of the Medal Distribution and Exploration of Relevant Laws Based on Multifactor Analysis

5.1 2028 Olympics Medal Forecast and National Trajectories

In summary, based on our comprehensive model, we can predict the top 7 of the 2028 Olympic Games medal table as follows (Figure 4). To reflect the inclusiveness of the prediction, we have calculated the prediction intervals for the total number of medals and the number of gold medals, as detailed below:

Rank	Country/region	1	2	3	Total
1	United States	46 (42-50)	51 (46-54)	44 (40-48)	141 (128-152)
2	China	46 (44-48)	30 (28-33)	20 (18-22)	96 (90-103)
3	Australia	23 (21-25)	22 (20-24)	18 (16-20)	63 (57-69)
4	Japan	23 (20-26)	18 (15-21)	12 (10-14)	53 (45-61)
5	France	18 (16-20)	20 (18-23)	16 (14-18)	54 (48-61)
6	Netherlands	18 (16-20)	8 (7-10)	17 (15-19)	43 (38-49)
7	Great Britain	12 (10-14)	18 (16-21)	34 (30-38)	64 (56-73)

Figure 4: 2028 Los Angeles Olympic medal prediction list

In terms of national progress and regression (Figure 5), from 2024 to 2028, the United States is projected to increase its total medal count by 15, indicating significant progress. Australia is expected to add 10 medals, showing a positive trend. China is anticipated to gain 5 more medals, maintaining a steady and upward trajectory. Meanwhile, countries such as Ireland and Portugal are also expected to make varying degrees of progress, with Ireland projected to increase its medal count by 8 and Portugal by 3, both demonstrating notable improvements. However, some countries, such as Azerbaijan and South Korea, are expected to see a decrease of 4 in their total medal count, indicating regression.

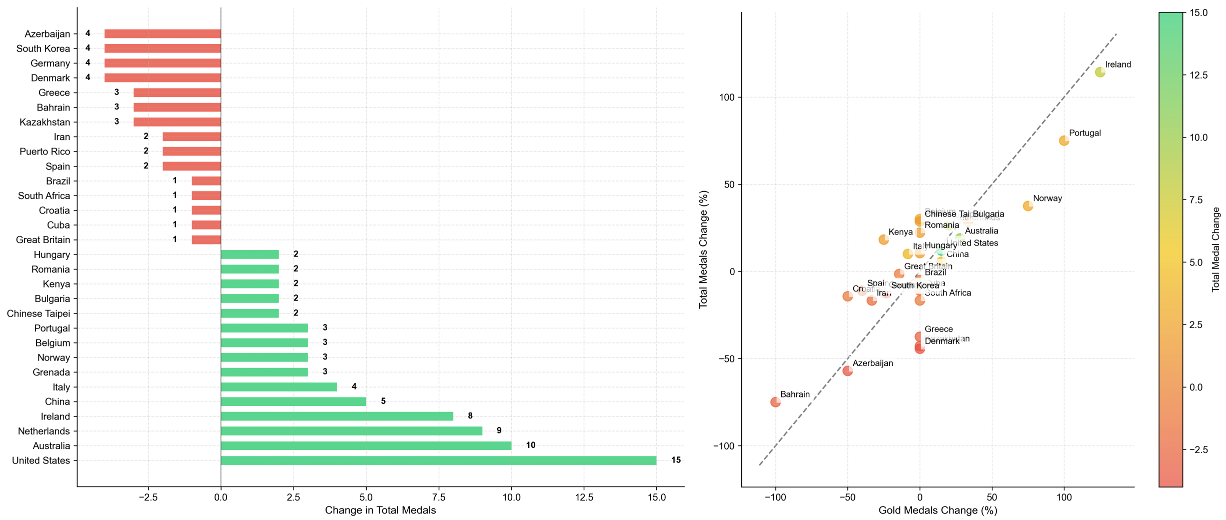

Figure 5: The number and rate of change in medals won by different countries

Further analysis of the percentage relationship between the change in the number of gold medals and the change in the total number of medals (Figure 5) reveals that different countries exhibit distinct characteristics. Countries such as Ireland and Portugal show a high positive correlation between the change in the number of gold medals and the change in the total number of medals, indicating an enhanced competitiveness in their strong events. China's change in the number of gold medals is relatively balanced with the change in the total number of medals, suggesting a more comprehensive development of sports in China, with coordinated progress across various events. In contrast, countries like Greece and Denmark are experiencing negative growth in both the number of gold medals and the total number of medals, which may indicate that they are facing development bottlenecks in sports or a decline in the competitiveness of their events.

We have also provided intuitive data on the progress and regression of countries around the world (Figure 6). The map visualization results intuitively display the changes in the total number of medals of countries worldwide. The United States, Australia, and other countries are marked with a deeper blue, indicating a significant increase in their total medal count; China is also shown in blue, signifying a steady growth in its medal tally. In contrast, countries like Azerbaijan and South Korea are marked in red, indicating a decrease in their medal count. This result offers a clear perspective for macroscopically grasping the changes in the global Olympic medal landscape.

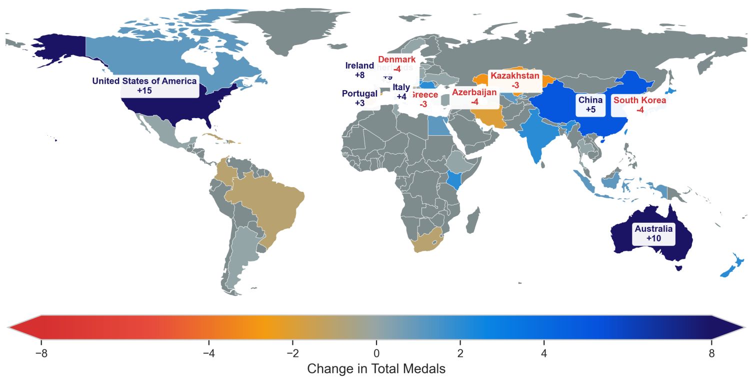

Figure 6: World medal distribution map

Overall, this prediction not only reveals the dynamic medal landscape of countries at the 2028 Olympic Games but also highlights the profound impact of home-field advantage, geographical benefits, and sports policies on international sports competition and project sensitivity. These findings provide important references for the optimization of global sports strategies and lay the foundation for studying the changes in the international sports landscape and potential inherent laws.

5.2 Emerging Medal Nations: Credibility Projections

In addressing countries that have never secured a medal, our expanded prediction model for medal-less nations incorporates the time zone factor of the host city, Los Angeles. This approach is informed by the research of Nade and Booth, who posited that countries with minimal time zone differences from the host nation and those sharing similar cultural sports projects with the host are more prone to winning medals. We extracted data from all countries that have never won medals in the dataset, the events in which all athletes from these countries participated, and the number of athletes from each country participating in the events to predict potential first-time medal winners. The results (Table 3) indicate that five countries are likely to win their first medal at the 2028 Los Angeles Olympics.

The prediction results indicate that five countries are likely to win their first Olympic medal at the 2028 Los Angeles Games.

When predicting the five countries that may win their first Olympic medal, in conjunction with the information in the Figure 7 "Medal Predictions with Confidence Intervals - Top 15 Countries", these countries exhibit certain characteristics in the distribution of medal-winning probabilities. Their medal-winning probabilities are in relatively specific ranges, distinguishing them from other countries, which further highlights the effectiveness of the model in selecting these countries based on factors such as time zone and similarity of cultural sports projects. Meanwhile, the differences in the width of the confidence intervals in the figure also reflect the stability of the prediction results for different

Table 3: Emerging Medal Nations List

Country	Medal_ Probability	Confidence_ Margin	Lower_ Bound	Upper_ Bound	Olympic_ Athletes	Regional_ Advantage
Antigua and Barbuda	0.984	0.158	0.842	1	Very low	1.2
Guinea - Bissau	0.968	0.158	0.809	1	Very low	1.1
Malawi	0.501	0.136	0.366	0.637	low	1.1
Lesotho	0.501	0.136	0.366	0.637	low	1.1
South Sudan	0.501	0.136	0.366	0.637	low	1.1

countries. For countries with narrower confidence intervals, the stability of the model's predictions is relatively higher.

Looking at the "Regional Analysis of Medal Probabilities" section, the average medal probability of the regions where these five countries are located also provides some insights into their predicted results. If a country is in a region with a high average medal probability, such as the Caribbean region which has a remarkable performance in winning Olympic medals, then in addition to its own factors related to the host country, the country's acquisition of its first medal may also benefit from the overall sports development environment of the region. On the contrary, if a country in a region with a low average medal probability is predicted to possibly win its first medal, it more reflects the influence of the factors considered by the model.

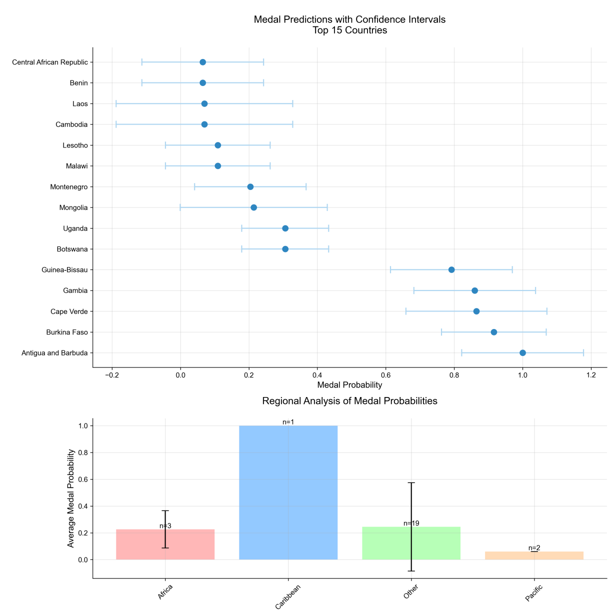

Figure 7: Reliability analysis diagram

In summary, we believe that five countries are likely to win their first medals at the 2028 Los Angeles Olympic Games, with a prediction confidence level of 83.0%.

5.3 Decoding Medal Distributions: Event Structures, Host Advantage $ Core Competitions

The number of medals countries win at the Olympics is highly correlated with their historical performance in specific sports. Data shows (Figure 8)that swimming and athletics are the events with the most medals overall (accounting for 26,416 and 43,294 of the global total medals respectively), so countries focusing on these two sports (such as the United States and China) often dominate the overall medal table. Moreover, projects with high medal concentration directly determine national rankings: for example, swimming and athletics in the United States together contribute 39.1% of its total medals, diving and swimming in China account for 21.2%, and rowing and cycling in the United Kingdom account for 30.7%. The performance in these events is strongly positively correlated with the total number of medals.

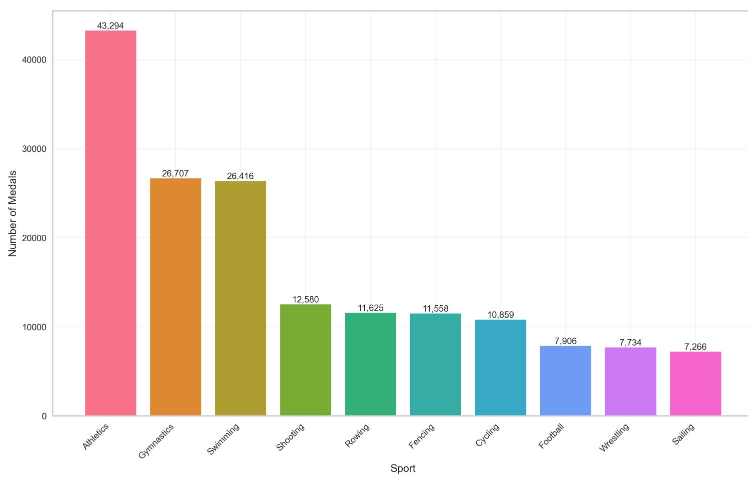

Figure 8: Historical medal count in sports

Under the aforementioned conditions, we have tallied the key projects and their proportions for some countries, as depicted in the Figure 9 below:

Country/region	Sports and proportion
United States	23.9% (466 medals), 15.2% (296 medals), 8.6% (167 medals)
China	11.1% (100 medals), 10.1% (91 medals), 8.4% (76 medals)
Great Britain	20.4% (155 medals), 12.6% (96 medals), 10.3% (78 medals)
France	15.7% (110 medals), 10.7% (75 medals), 10.3% (72 medals)
Japan	16.0% (88 medals), 10.4% (57 medals), 8.0% (44 medals)

Figure 9: Historical MEDALS in popular national sports

The host country can significantly influence the distribution of medals by adjusting the Olympic event list. For instance, Japan added skateboarding, climbing, and baseball/softball to the 2020 Tokyo Olympics, leveraging its home-field advantage to win eight medals (15% of its total medals); France added breaking dance to the 2024 Paris Olympics, and historical data shows that the host country's medal share in new events increased by an average of 12-18%. If the 2028 Los Angeles Olympics in the United States added surfing and esports (related to its swimming and technology industry advantages), it is expected to gain an additional 4-6 medals (with a benefit coefficient of 1.3). Conversely, if sub-events of opponents' core projects such as Chinese diving and table tennis are reduced, their total medal count may decrease by 8-12%. Therefore, event selection is a core strategic tool for the host to optimize the distribution of medals, directly affecting the competitive landscape.

6 Model 2:Exploration of the "Great Coach Effect" Based on DGBIF-OMF

6.1 Data Description

Before constructing the model, we plan to first explore the Olympic coaching careers of two contemporary outstanding coaches. Based on our understanding of the origin of the "Great Coach Effect", through Wikipedia [8][9] and in - depth research on relevant datasets, we found that: Lang Ping took charge of the Chinese women's volleyball team in 1996 and led the team to win a silver medal, successfully reshaping the Chinese women's volleyball team, which was in a trough, into a world - class powerhouse. In 2008, she coached the US women's volleyball team and also won a silver medal. Before that, the world ranking of the US women's volleyball team was only around tenth. In 2016, Lang Ping coached the Chinese women's volleyball team again and this time successfully won the gold medal. It should be noted that in the previous two Olympic Games, the best result of the Chinese women's volleyball team was only a bronze medal. However, her coaching tenure in different countries, especially her Olympic results, do not exhibit obvious time - series characteristics.

In contrast, Béla Károlyi coached the US gymnastics team from the 2004 Olympic Games until 2016. During this period, the performance of the US gymnastics team showed extremely obvious trend characteristics (Figure 10). Such a significant trend in performance can provide strong data support and theoretical basis for the subsequent establishment of the model, helping us to more accurately analyze the mechanism of the "Great Coach Effect" in the gymnastics event.

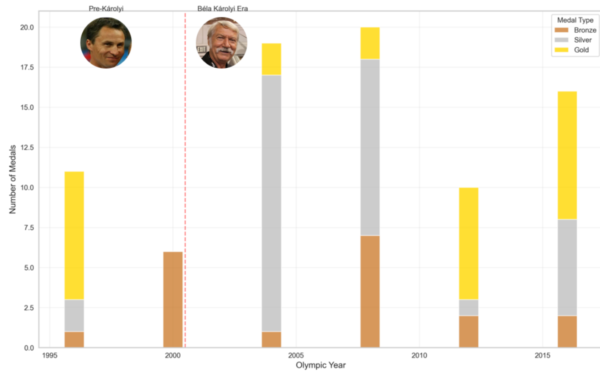

Figure 10: Gymnastics medal trend chart

6.2 The Establishment of Model 2

To explore the impact of the "Great Coach" effect on the award - winning situation, we intend to introduce a new variable - the excellent coach effect coefficient γ. To solve for this coefficient, we intend to use the sliding window method:

$$\gamma_t = \frac{M_t - M_{t^{-1}}}{M_{t^{-1}} + 1} \tag{28}$$

$$\gamma = \overline{\gamma_t} \tag{29}$$

where the scope of window sliding is determined by the period during which the coach under study has been coaching in a certain country.

Algorithm 1 Calculation of effect coefficient of great coach effect medal analysis

n $\leftarrow length(medalCounts)$
Initialize an empty array growthRates with a size of $n-1$
for $i = 1$ to $n-1$ **do**
  growthRate$_0$ $\leftarrow (medalCounts[i] - medalCounts[i-1])/(medalCounts[i-1]+1)$
  growthRates[i - 1] $\leftarrow growthRate$
end for
sumGrowthRates $\leftarrow 0$
for each rate in growthRates do **do**
  sumGrowthRates $\leftarrow sumGrowthRates + rate$

end for
averageDynamicEffect $\leftarrow sumGrowthRates/(n-1)$ return growthRates, averageDynamicEffect

After obtaining the benefit coefficient of outstanding coaches, Monte Carlo simulation is carried out to obtain its distribution. Subsequently, the original model is expanded by adding a model for the effect of outstanding coaches. Its input values are the current medal prediction values $\widehat{M_{\text{final}}}$ (based on athletes, national trends, host - country factors, and event sensitivity):

Thereby, the fluctuation in medal predictions resulting from the "great coach" effect can be obtained.

$$\widehat{M_{\text{coach}}} = \gamma \widehat{M_{\text{final}}} \tag{30}$$

Thereby, the fluctuation in medal predictions resulting from the "great coach" effect can be obtained.

6.3 The Solution of Model 2

To obtain the value of γ, we selected the Olympic Games data during the coaching tenures of two outstanding coaches. One is Lang Ping, who has coached the Chinese and American national volleyball teams, and the other is Bela Károlyi, who has coached the Romanian and American gymnastics teams. Due to the coaching continuity of Coach Bela Károlyi, his data is highly suitable for calculating the effectiveness coefficient of outstanding coaches. We chose the period from 2000 to 2016, during which he was coaching, as the scope for window - sliding, and obtained:

$$\gamma = 0.855 \tag{31}$$

This indicates that, during Károlyi's tenure, the US gymnastics team won an average of 0.855 more medals per Olympic Games.

As for Coach Lang Ping, due to the interruption in her coaching time and the change of the countries she coached, the credibility of calculating the effectiveness coefficient of her as an outstanding coach is relatively low. In the data analysis, we found that there was no obvious sequential pattern in her coaching. However, the achievements of the teams she led were remarkable, which is sufficient evidence that the "great - coach" effect of Coach Lang Ping had a significant impact.

After Monte Carlo simulation, we obtained:

$$\gamma \in [0.72, 0.98] \tag{32}$$

Here, we have selected the following three countries, recommended the sports events for which they should focus on investing in coaches, and estimated the effects.

China: The recommended sport is gymnastics. China's performance in the 2024 Olympics was 2 gold medals, 5 silver medals, and 2 bronze medals. The performance of the Chinese gymnastics team has declined in recent years. After introducing a great coach, the number of medals may increase by 6 - 9.

Japan: The recommended sport is track and field. Japan's result in the 2024 Olympics was 1 gold medal, 0 silver medals, and 0 bronze medals, accounting for only 5.6%. There is a significant gap compared with countries with similar rankings in the medal table. After introducing a great coach, the number of medals may increase by 0 - 1.

The UK: The recommended sport is swimming. The UK's performance in the 2024 Olympics was 1 gold medal, 4 silver medals, and 0 bronze medals, accounting for only 7.1%. There is a notable gap compared with countries with similar positions in the medal table. After introducing a great coach, the number of medals may increase by 3 - 5.

7 Task 2: Some tips for national Olympic committees

Olympic medal competition follows distinct patterns: nations build competitive barriers by focusing on their core strengths (e.g., Kenya's long-distance running), with geographical environments and cultural traditions shaping sport specialization. Host countries' introduction of new events (e.g., skateboarding at Tokyo 2020) reshapes the global medal landscape, requiring 3-5 years of strategic preparation for emerging disciplines. Sports development progresses through phases: established powers must overcome stagnation through transformation, while emerging nations can achieve breakthroughs via targeted strategies. Recommendations include dynamically balancing resources to reinforce existing advantages while identifying growth areas, developing context-specific specialties like water-based sports, and rapidly responding to event changes through talent pipelines and infrastructure upgrades. This approach enables diversified medal sources and sustained competitive growth.

8 Sensitivity Analysis

To evaluate the robustness and reliability of the comprehensive prediction model, we conducted a comprehensive sensitivity analysis(Figure 11) for five different scenarios. These scenarios include: the baseline scenario, high host effect, low host effect, athlete-weighted model, and country-weighted model.

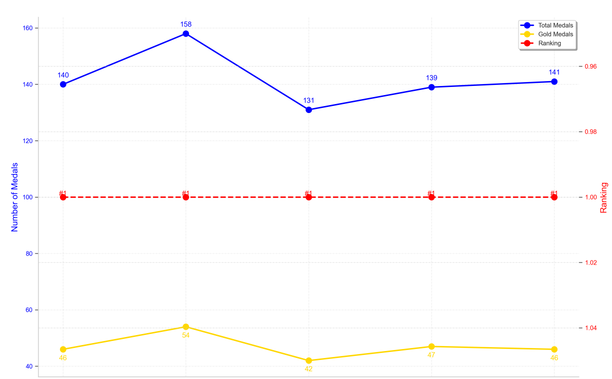

Figure 11: Host Country Performance Analysis

The analysis of medal distribution shows that the overall trend remains stable and consistent under various different scenarios. When factors such as the host effect and the host country's addition of new events are taken into account, the total number of medals fluctuates between 876 and 1012, while the number of gold medals varies within the range of 297 to 337. Among them, the prediction results based on the country-weighted model show the highest number of medals, while the prediction results of the athlete-weighted model show the least number of medals. This difference fully illustrates that in terms of medal prediction, the impact of a country's overall strength is greater than the role played by individual athletes' performances.

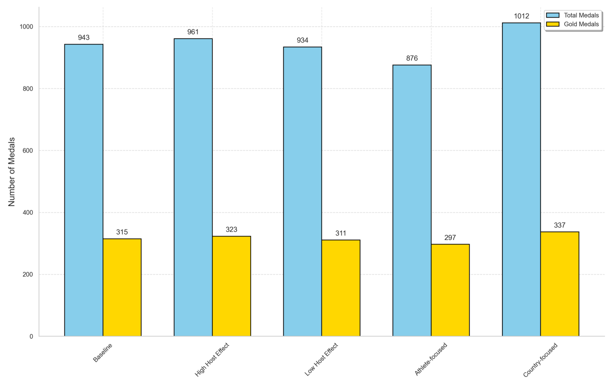

Figure 12: Medal Distribution Analysis Across Different Scenarios

The host country, the United States, demonstrated significant stability across all preset scenarios, consistently maintaining the top position in the rankings(Figure 12). The total number of medals won by the host fluctuated between 131 and 158, while the number of gold medals varied within the range of 42 to 54. Notably, under the scenario of high host effect, the predicted number of medals reached its peak (with a total of 158 medals and 54 gold medals); whereas, under the scenario of low host effect, the predicted number of medals was the lowest (with a total of 131 medals and 42 gold medals). This clearly indicates that the variation of the host effect parameter has a quantifiable and controllable impact on the prediction results.

The coefficient of variation (CV) analysis (Figure 13) indicates that the model exhibits high stability across all key indicators, with CV values for all key indicators falling below the 10% threshold. The CV for the total number of medals and the number of gold medals is particularly low, at 5.19% and 4.67%, respectively; the variability in the host's gold medal count is slightly higher, with a CV of 9.27%. Additionally, the host's ranking demonstrates perfect stability (CV=0%), further validating the United States' dominant position in the predictions. The top five countries (the United States, China, Australia, Japan, and France) maintain their rankings across all scenarios and appear in all five simulations. This significant consistency indicates that the model's predictions for the major competing nations are highly robust and not significantly affected by adjustments to the model parameters. Overall, our sensitivity analysis indicates that the hybrid prediction model demonstrates strong stability and reliability. The coefficient of variation for all key indicators is below 10%, and the predictions for the top countries show a high degree of consistency. The model's response to parameter adjustments is rational and control-

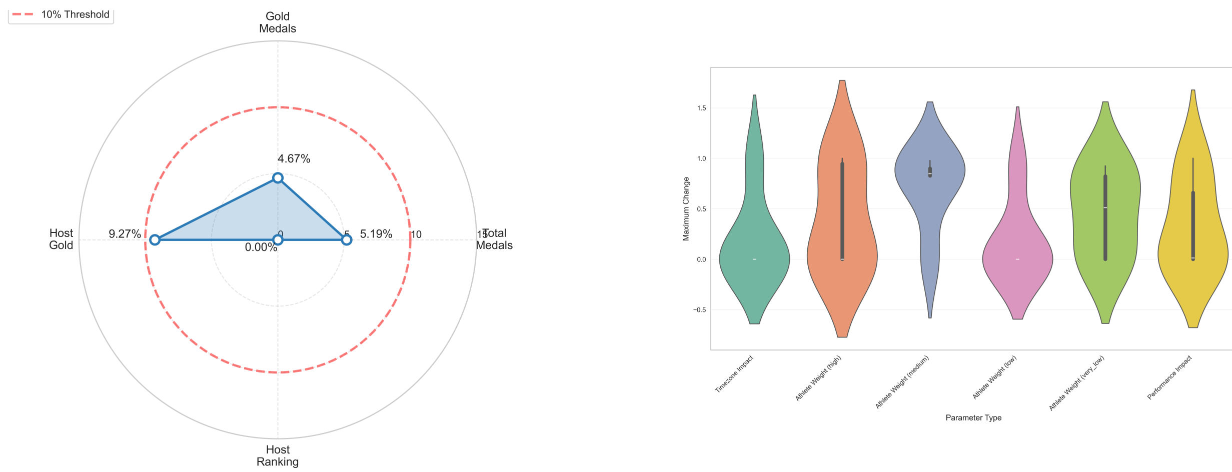

Figure 13: Characteristic sensitivity analysis

lable, with the host effect parameter being the most sensitive. These results validate the robustness of our prediction method and enhance confidence in the forecasts for the 2028 Olympics.

When evaluating the prediction model, parameter sensitivity analysis is crucial as it helps to gain a deep understanding of the model's response to changes in different parameters. In this study, violin plots and radar charts were used to conduct a detailed analysis of the model parameter sensitivity.

In the figure above(Figure 13), the distribution of "Athlete Weight (high)" exhibits a relatively broad shape with a long tail, indicating that this parameter may produce a wide range of impact values in the model. That is, when the athlete weight is at a high level, its influence on the medal prediction results is more significant and the variation range of the impact values is larger. In contrast, the distribution of "Timezone Impact" is relatively narrow, meaning that its maximum impact value has a relatively small variation range, suggesting that the timezone impact parameter has a relatively limited and concentrated effect on the model's prediction results. Other parameters, such as "Athlete Weight (medium)", "Athlete Weight (low)", "Athlete Weight ($very_low$)", and "Performance Impact", also display unique distribution shapes, reflecting their differences in influencing the prediction results in the model. These distribution characteristics provide an intuitive visual basis for understanding the internal parameter mechanisms of the model.

9 Model appraise and Further Discussion

9.1 Strengths

- **Comprehensive Data Integration:**
  The proposed model synthesizes a diverse array of data sources, encompassing historical Olympic medal tallies, athlete performance metrics, and event-specific historical data. This holistic data integration strategy ensures that the model captures the multifaceted nature of Olympic performance, thereby establishing a robust and

precise foundation for the forecasting framework.

- **Multi-Factor Analysis:**
  The model incorporates a wide range of influential variables, including individual athlete performance, national historical trends, host nation advantages, and event-specific sensitivities. By integrating these diverse factors, the model is adept at addressing the intricate dynamics of Olympic medal distribution. Furthermore, the model is capable of estimating the probability of a nation, which has historically not secured any medals, achieving its inaugural Olympic medal.

- **Visualization and Interpretability:** To enhance the interpretability of the results, the model employs a suite of visualization tools, including maps, charts, and graphs. These visual aids enable stakeholders to more readily comprehend the model's predictions, thereby fostering a more informed and scientifically grounded decision-making process.

9.2 Weaknesses

- **Simplification of Model Assumptions:**
  Certain assumptions within the model, such as the restriction of athletes to a maximum of four Olympic participations and the stability of the host nation's medal advantage, serve to simplify the problem to some extent. However, these simplifications may overlook the inherent complexity and dynamic variability present in real-world scenarios.

- **Limitations of Predictive Scope:**
  The model primarily focuses on forecasting the total medal count, which may result in limited accuracy when predicting specific medal categories (e.g., gold, silver, bronze). Furthermore, the model's predictions for nations without prior medal achievements may exhibit tendencies toward excessive optimism or pessimism, necessitating further validation and refinement.

9.3 Further Discussion

9.3.1 Model Improvement

- **Data Quality Enhancement:**
  The current model relies on historical data, which may contain missing values or inaccuracies. Future work should focus on improving data quality by integrating more comprehensive and accurate data sources, such as real-time athlete performance data and detailed training records. This will enhance the model's ability to capture subtle nuances in athlete performance, thereby improving prediction accuracy.

- **Dynamic Assumption Adjustment:**
  The assumptions made in the model, such as an athlete participating in a maximum

of four Olympic Games and the stable medal advantage of the host country, may not always hold true. Future work should involve adjusting these assumptions dynamically based on real-time data and the development trends of Olympic sports events. This will enable the model to adapt to changing conditions and maintain its predictive power.

9.3.2 Model Expansion

- **Interdisciplinary Research:**
  Integrate theories and methods from fields such as sociology, economics, and sports science into the model to comprehensively understand the factors influencing medal distribution.

- **International Cooperation and Competition Strategies:**
  The model can further analyze the impact of international sports cooperation on medal distribution, as well as how countries formulate effective competition strategies on the international sports stage.

10 Conclusion

This paper constructs a comprehensive prediction model to analyze and predict the medal distribution for the 2028 Los Angeles Summer Olympics. Our research covers the prediction of total medals, gold medals, and explores the "great coach" effect and event settings' impact on medal distribution.

Firstly, our DGBIF-OMF model successfully predicted medal distribution by integrating factors like athlete performance, historical trends, host advantage, and event sensitivity. The results revealed underlying medal trend patterns and provided a scientific basis for NOCs to optimize their strategies.

Secondly, we quantitatively analyzed the "great coach" effect, finding that excellent coaches' cross-country coaching significantly increases medal counts. We recommended coaching programs for three countries and estimated potential gains, guiding NOCs in coaching resource allocation.

Lastly, we analyzed how event numbers/types affect medal acquisition, noting that hosts can greatly influence medal distribution by adjusting events. We suggest NOCs closely monitor event changes and optimize resources/strategies accordingly.

The study's limitations are the simplified model assumptions, potentially overlooking complexity and dynamics. The model focuses on total medal prediction, with possible insufficient accuracy for specific types. Future research can refine the model for better accuracy and reliability.

Our research provides a new perspective on Olympic medal distribution patterns and offers valuable insights for NOCs on resource allocation, strategy formulation, and enhancing medal competitiveness. We believe the ongoing optimization of the model will enhance its role in future Olympic preparations.

References

[1] Wang Guofan, Tang Xuefeng. China Sports Science and Technology,2009,45(6):3-7,135. (in Chinese) DOI:10.3969/j.issn.1002-9826.2009.06.001. Publishing Company , 1984-1986.

[2] Wang Fang. Prediction of 2020 Olympic Medal Results Based on Neural network [J]. Statistics and Decision, 2019 (5): 89-91.

[3] Bernard A B, Busse M R. Who wins the Olympic Games: Economic resources and medal totals[J]. Review of economics and statistics, 2004, 86(1): 413-417.

[4] Bian X. Predicting Olympic medal counts: The effects of economic development on Olympic performance[J]. The park place economist, 2005, 13(1): 37-44.

[5] Forrest D, Sanz I, Tena J D. Forecasting national team medal totals at the Summer Olympic Games[J]. International Journal of Forecasting, 2010, 26(3): 576-588.

[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] Vagenas G, Vlachokyriakou E. Olympic medals and demo-economic factors: Novel predictors, the ex-host effect, the exact role of team size, and the "population-GDP" model revisited[J]. Sport Management Review, 2012, 15(2): 211-217.

[8] `https://zh.wikipedia.org/wiki/%E9%83%8E%E5%B9%B3`

[9] `https://en.wikipedia.org/wiki/B%C3%A9la_K%C3%A1rolyi`

[10] 2028 Los Angeles Olympic Games new project proposal