World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3100

THE SPATIO-TEMPORAL EVOLUTION OF OLYMPIC MEDALS BASED ON MULTILEVEL REGRESSION ANALYSIS AND MULTIDIMENSIONAL FEATURE COUPLING

YuHang Xiao
School of Science, Beijing University of Civil Engineering and Architecture, Beijing 102600, China.

Abstract: Olympic medal performance is a measure of a country's competitive sport achievements but also largely shaped by various socioeconomic aspects. This study attempts to comprehensively investigate the spatiotemporal development of Olympic performance and to build a highly accurate hierarchical forecasting model. Firstly, using data from the 2008-2024 Olympic Games (five editions) after data cleaning and feature engineering it can be observed an extremely positive correlation between size of the games and total number of medals with correlation coefficient. The exploratory analysis suggests that GDP is the main socio-economic factor determining a nation's medal competitiveness and its relationship with the number of medals has significant structural differences. It can be seen that for high-income countries, the conversion efficiency of economic resources into competitive advantages is best. This research devised separate forecast systems for nations on varying levels of competition; it used a share-smoothing prediction method on top-performing countries so grasp long-term trend; added up weight least-squares shares regression plus reliability shrinkage system for mid-ranking countries; and innovatively developed a two-step prediction framework integrating logistic regression and random forests models for those nations without any previous medal record. Forecast results suggest at the 2028 Los Angeles Olympics USA will get 150 medals because they are the host nation while China UK and France which are all powerhouses will stay in the lead positions. With the coupling of multiple features, this paper gives scientific evidence for people to know about the change in Olympic performance.
Keywords: Medal distribution attribution; Hierarchical forecasting framework; Socioeconomic determinants

1 INTRODUCTION

Olympic medal standings are competitive in the context of sports as a small-scale representation of global comprehensive national strength. Researching what leads to good performance by countries at the Olympics helps us understand more about how athletic skill changes over time and space, and also gives nations numbers they can use when making careful plans for getting ready. In terms of prediction model research, the current problems mainly lie in the quantification of non-linear effects of socioeconomic factors on the output of medals, and also in the different stabilities of countries' performances at various levels. Previous studies mainly used single regression models or simple linear extrapolation which ignored the role of economic development level in affecting the efficiency of turning out medals, it did not have proper probabilistic methods to forecast if any nation will make "zero-breakthrough" i.e., achieving its first ever medal. Innovation here is a performance - stratified coupled forecasting framework: Using share-smoothing approaches together with a two-step machine learning architecture, it allows specific modeling for stable players, fluctuating players, and potential breakthroughs. Overall approach for research would be as follows: firstly, by preprocessing and feature engineering data from previous 5 Olympic Games we evaluate features like GDP, population and host nation effect; secondly through pearson correlation analysis and analysis of variance we break down the distribution of medal winning across all sport categories and differing economic development levels; following this, competitors were split up into three groupings depending upon competitiveness level and applying trend fitting using linear trends, weighted slope estimation method and logistic regression- random forest combination model projected the trend of medal distributions until 2028; finally assessing the strength of such model dealing with complicated uneven information integrated with predicted outcomes[1-3].

2 MODEL

2.1 Data Preprocessing

In this question, we got quite a few raw data sets which contain information about Olympic medals and try to find out if there is some connection between these factors and the medal. The data is from https://mcmicm.org.cn/. In order to guarantee the correctness and consistency of the data so offer a dependable foundation for the following analysis and modeling work, we did thorough processing on the data [4].

2.1.1 Data cleaning

In order to carry out the following analysis, we first cleaned the data with Python in order to get rid of any error which might influence the accuracy of the results. Specific steps are as follows:
missing value processing: when there were missing or incomplete records, we chose suitable methods for filling or deleting them. Original data choose to delete Russia's data, as Russia could not compete in 2024 and had zero golds and

medals for 2024. such data cannot be used for training the next stage machine learning models, so it is necessary to remove these from the dataset. It can avoid the bias caused by the missing values on the analysis results effectively and guarantee the reliability of the analysis results[5].

Duplicate value processing: take each record and compare it with other records in the dataset based on their key field information. If all of a record's key field information matches exactly that of another record, then these two records are considered duplicates. After finding out which records were repeated, we only kept one record from every group of duplicates, and removed the rest. By means of this operation, only unique and valid records remain within the data set, creating an excellent database for later analyses and modeling work [6].

2.1.2 Data filtering

In order to focus on data related to the last five Olympic Games, we screened records from the original dataset for the years 2008, 2012, 2016, 2020, and 2024. This screening step allowed us to limit the timeframe of the analysis and ensure that the subsequent study was based only on data from the last five Olympic Games. Subsequently, the screened data are grouped according to countries and regions as well as years, and the Gold and Medal of each country and region in each Olympic Games are calculated, through the above processing, the specific award-winning situation of each country and region in each Olympic Games can be clearly counted, which can provide basic data support for the subsequent analysis[7-8].

In the process of data screening, this paper further adopts a rule-based approach to streamline the data in order to retain records that are of significant value for subsequent analysis or modeling. Specifically, we set the following screening condition: if a country or region has won more than 50 gold medals in a certain Olympics, it is considered to have a high level of competition in that Olympics. Based on this rule, countries such as China and the United States, which have won more than 50 gold medals in some years, were selected from the data and meet the above definition of "high level" countries.

Similar to the definition of "high level" countries, we also define "medium level" countries, which are categorized into two groups, namely, countries that have won medals in two to five tournaments and countries that have won medals in only one tournament. Based on this rule, we filtered the data to include countries such as Canada, which met the above definition of "medium level".

We also filtered out countries that have never won a medal to predict which countries will win their first medal at the next Olympics[9-10].

2.1.3 Feature creation

Feature creation is the first stage in the process of creating features. Creating new feature by using the current data Creation of features: Based on the original data, we think if a country is an organizing country will affect the number of medals, so based on the information of whether a country is an organizing country, we design a new feature to indicate whether each country is an organizing country in every Olympics.

This new feature can help the model better capture the impact of the host country on the number of medals, as host countries usually perform better in the medal table, which can be related to factors such as home advantage, spectator support, better training facilities, and so on.

2.1.4 Data visualization

In order to understand the distribution and characteristics of the data more intuitively, we worked on data visualization using Python.

We use the pandas library to process data and matplotlib and geopandas libraries to create maps. By creating the map shown below, we can see how GDP varies across different countries. This will allow us to see which countries are doing well economically and which countries could benefit from more investment in sports. And also the GDP is also a factor used when trying to predict the number of medals for a certain country during the Olympics since economic strength may affect how much a nation invests into sports and the training conditions for their athletes.

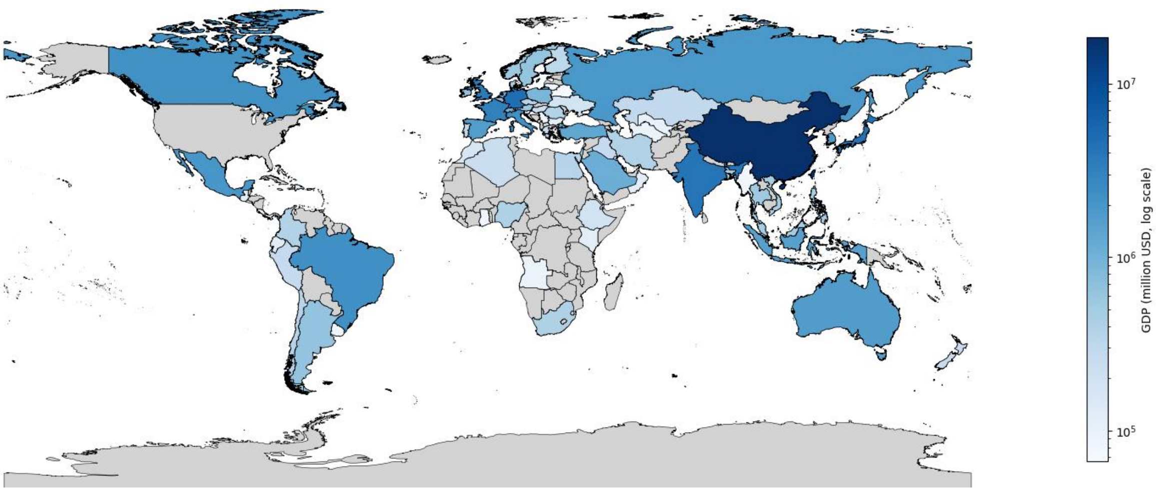

Figure 1 Global GDP Distribution by Country (Self-Drawn)
Source: http://bzdt.ch.mnr.gov.cn/

Global GDP distribution by country is shown in Figure 1.

2.2 Exploratory Tests

2.2.1 Relationship between event scale and medal output

We applied the use of Pearson Correlation Analysis and scatter plots to analyze event count in relation to medal count for different sports. The Pearson's correlation coefficient is a statistic which indicates the strength and direction of linear relationship between two variables, it can take values from -1 to 1 The coefficient is calculated as follows:

$$\mathrm{r} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{1}$$

where x_i and y_i represent the observed values of the two variables, and $\bar{x}$ and $\bar{y}$ are their means. Its value is used to measure the degree of linear correlation between two numerical variables, and its value ranges from -1 to 1. The closer the absolute value is to 1, the stronger the correlation is, while 0 means no linear correlation.

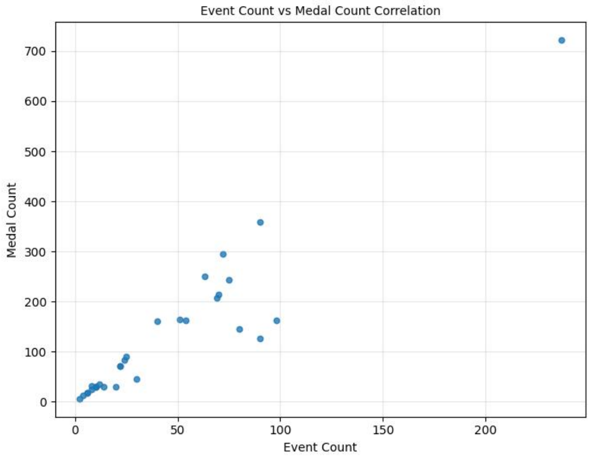

Figure 2 Relationship between Event Count and Medal Count

Relationship between event count and medal count is shown in Figure 2.

Scatterplot shows an apparent positive relationship between event count and medal count. With data on 30 sports, we obtain a Pearson correlation coefficient of 0.94, implying there is a strong positive linear correlation The corresponding p-value equals 7.24×10^{-15}, so it's statistically significant Therefore, the more events in a sport, the more opportunities for medals, so the total number of medals will be greater.

2.2.2 Socioeconomic determinants of Olympic medal performance

In light of the above analysis on event scale, we go on to look at the link between socio-economic aspects and Olympic medal results. We pay particular attention to GDP and population size, which are often considered to be basic determinants of a country's ability in sports events. To reduce the effect of the different levels of scale as well as reduce heteroscedasticity, both the GDP and the population are log-transformed before being analyzed.

(1) Multivariate Correlation Structure

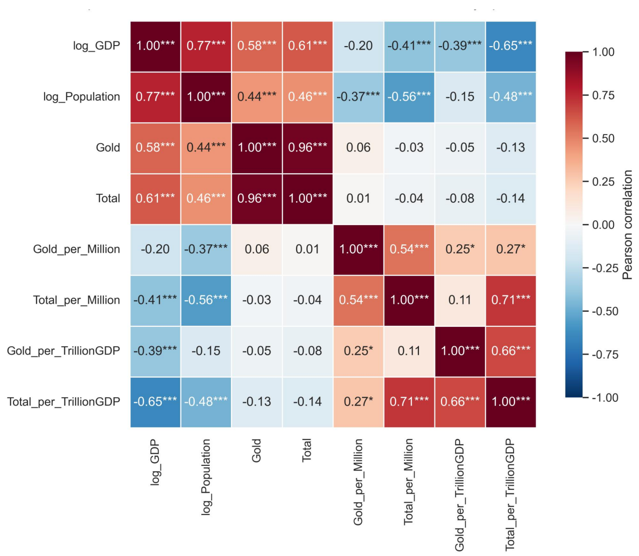

Figure 3 Expanded Correlation Matrix of Socioeconomic Variables and Olympic Performance

Figure 3: is a more broad correlation matrix which contains economics, demographics, medal number and efficiency metrics. Results show an obvious structural pattern, economic forces take the lead position in shaping up Olympic results. Log(GDP) has moderate positive correlations with gold medals and total medals, suggesting that countries with larger economies tend to do better overall in Olympic performance than other countries. On the contrary, log(population) has very little connection with the quantity of received medals, its coefficient for gold medal was just 0.44 and even less, at only 0.46 for all medals. So population size can give you more people to find talent among but it's not as good at helping you win compared to having money and resources.

Additionally, incorporating efficiency metrics like medals per million people and medals per unit GDP also displays an inverse trend with regards to GDP and population. These show that smaller or poorer countries can do better in changing their resources into medals than bigger and richer nations even if the latter actually get more total number of medals. To put all things together, it is evident from how the correlations work that GDP affects Olympic accomplishment more than does population size. It follows then, that economic investment as well as infrastructure and supporting organizations are important when it comes to producing elite athletes.

(2) Scatter Analysis and Structural Heterogeneity

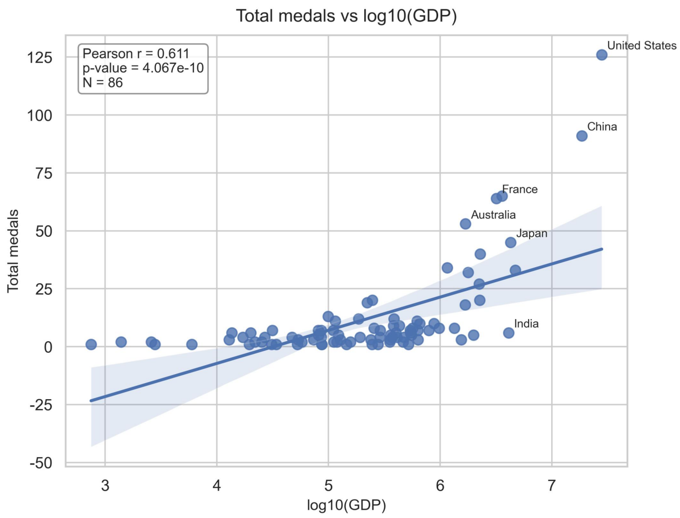

Figure 4 Relationship between Economic Strength and Olympic Medal Counts (log(GDP) vs Total Medals)

In order to further explore the connection between economic strength and medal performance, Figure 4 presents a scatter diagram that plots total medals against log(GDP), with an appropriate fitted regression line. The result is positive, with a p value below 0.001 and pearson correlation coefficient of 0.611 which means it's statistically significant This indicates that countries with more money generally win more medals although there are big differences from one country to another. High-income nations like the United States and China consistently top the medal count charts and sit above the regression trendline showing their powerful and consistent competitive edge. However on the other hand there are also some larger countries who have very less development economically like India these nations fall beneath the predicted trend - line; this suggests just because you have millions of people does not mean automatic success.

In addition, there are many countries at the medium level of income that have a big range of medals won. So it shows more things like national sport policy, sports training system and sports traditions also have great effect on performance. Thus GDP is the main explanation but can't explain all Olympic results variation so economic power vs medal tally link relies upon lots of different interlinked factors.

(3) Stratified Analysis by GDP Level

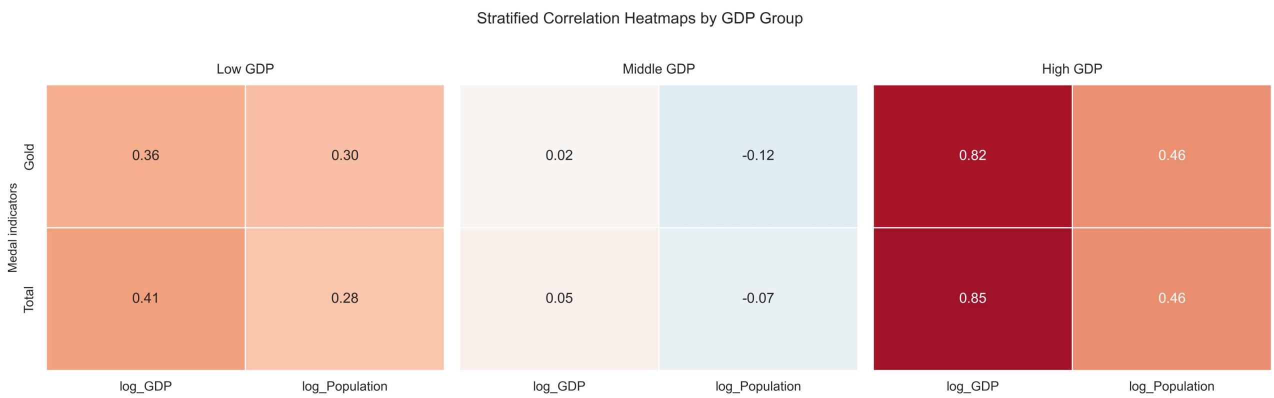

Figure 5 Stratified Correlation Heatmaps between Economic Indicators and Olympic Medal Counts across GDP Groups

In order to further explore the structural heterogeneity of the relationship between socioeconomic factors and medal performance, countries are categorized into three groups according to their GDP levels: low, middle, and high. From Figure 5's stratified correlation heatmap we see that how strongly correlated GDP is with medal count differs quite a bit from one group to another. In high-GDP countries, the correlation between GDP and medals counts is extremely strong, showing that economic resources can be turned into competitive benefits. These countries usually have sound sport systems, developed training infrastructure and good institutions which enable them to turn around their economic advantages into successes in winning medals.

But for countries with medium GDP it is a weak correlation and sometimes not even existent. It means other economic elements can't solely explain its success At this stage, policies' effectiveness varies, resources are distributed differently, as well as strategic investments on certain types of sports. Low GDP nations have a positive correlation as well but it is much weaker than in high GDP ones, which indicates that improving economy is still helpful in improving at sport although the effects is less since they lack a lot of resources.

These results indicate that the impact of GDP on Olympics isn't uniform, but rather varies according to where the nation stands economically. Thus, it's non-linear relationship and is very different between various groups of countries.

(4) Overall Interpretation

From the above analyses, it can be concluded that GDP as an indicator of national strength is a major factor affecting Olympic medal counts; population size has only a small and indirect impact. The connection between GDP and medals is both meaningful and structural, and changes across varying levels of development. And also these efficiency disparities suggest there are certain countries which can surpass their own economic boundaries by focusing on or investing heavily in some certain sports. From these insights we have a better grasp of how socioeconomic variables effect olympic results and it provides us with a factual basis for our upcoming modeling and forecasting work.

2.2.3 Distributional characteristics of Olympic medal events across sports

To investigate the way opportunities for achieving an Olympic medal differ across sports, we carried out a descriptive and inferential analysis of sports using de-duplicated records from medal events. To go into more detail: duplicate athlete-level medals entries were removed on the basis of composite Year+NOC+Sport+Event+Medal key; thus what was left behind are only true medal event observations rather than repeated athletes due to team competition record duplication This approach improves comparison validity across different type of sport, prevents over-count within teams. We then calculate the total number of medal events for every sport and show it in Figure 6. It is quite uneven. Athletics leads with 3292 medal events, followed by Swimming at 1890, Wrestling at 1440, Boxing at 1048, Gymnastics at 974. Shooting has 943, Rowing has 867, Cycling has 740, Fencing has 727, Weightlifting has 717. That means that there are few sports which have a lot of chances to win medals.

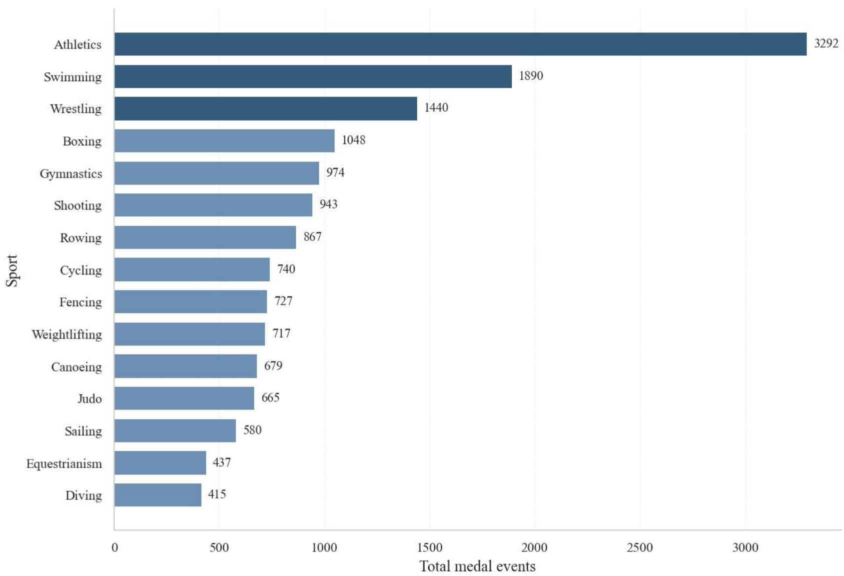

Figure 6 Total Number of Medal Events by Sport

This type of distribution will greatly affect the plan for national medals. Take athletics and swimming for instance, these types of sports have different kinds of events, so winning different sorts of medals in them, thus they add up more than just one sort of medal to an individual or a team's total. Those with few event categories can provide fewer opportunities to win medals: Therefore, it's more likely that maintaining competitiveness in large-volume sports rather than being good at small-volume ones will affect how well a nation does in terms of its overall medal count.

In order to explore the heterogeneity of sports more thoroughly, we created boxplots based on medal-event counts at the country-year-sport level, which can be seen in Figure 7. The results are quite different when it comes to central tendency and how much they vary. As for the average grouped medal-event count, Swimming tops the list with an average of 5.05 medal events per unit of country–year–sport followed by Athletics (4.40), Gymnastics (4.35), Diving (3.12) and Wrestling (3.07). Other sports with relatively large group means are Fencing (2.98), Canoeing (2.79), Table

Tennis (2.69), and Cycling (2.66). In some of these sports the mean exceeds the median and has very long upper tails implying a positive skew and suggesting that there is only a few countries that achieve outstanding performance in these disciplines.

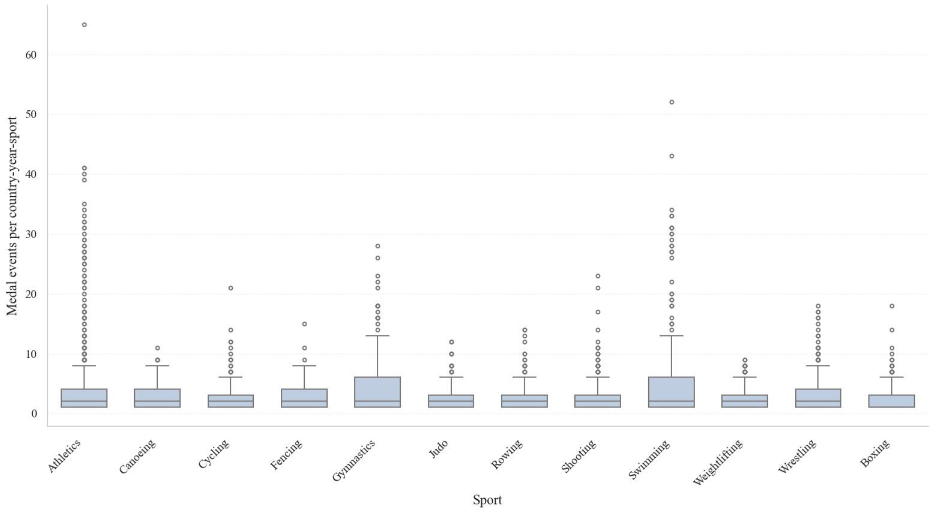

Figure 7 Distribution of Medal Events across Sports

To test if the above differences are statistically significant, we ran a one way ANOVA over sports. There was an extremely significant difference in group medal-event numbers between sports. Due to the boxplots showing signs of skewness and differing dispersion among groups, we also ran a Kruskal-Wallis test as a nonparametric robustness check. This was also highly significant, so we can confirm that this variation by sport is not due to distributional assumptions.
We further examined the size of this sport impact with effect-size measurements. Estimated eta squared equals 0.0906, omega squared is 0.0831 which means a little but significant influence from the type of sport for the distribution of medals-events. In other words, cross-sport variation accounts for a non-negligible part of the observed differences in grouped medal-event counts, yet within-sport variability stays considerable.
Figure 8, as further interpretive clarification, displays the average medal-event counts for each country-year-sport unit and its associated 95 percent confidence intervals for the top sport. Figure is showing Swimming ranks first in total number of medals as well as grouped average number of Medal-Events; followed by Athletics and Gymnastics: which implies these three have significant structural position regarding Olympic Medals.

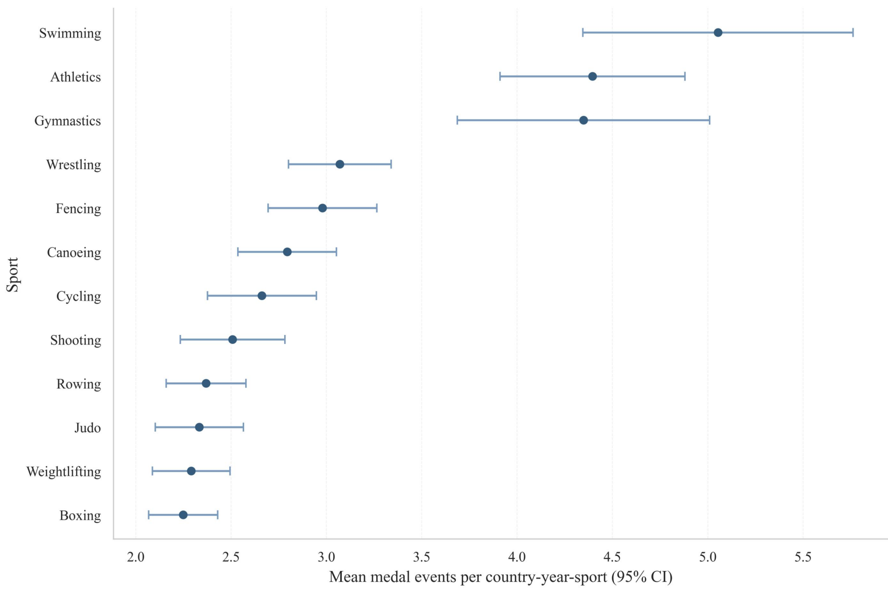

Figure 8 Mean Medal Events with 95% Confidence Intervals

On the whole, the analysis shows that Olympic medal events differ by sport, and these differences are statistically significant when tested using both parametric and nonparametric tests. Athletica is at the top for swimming in terms of total medal-events, followed closely by Swimming, Athletics, and Gymnastics which also have relatively large grouped-medal-event quantities at the country-year-sport level. These figures show there’s quantitative proof that you stand more chance of getting a medal doing certain sorts of sport than other types, so not all sports offer equal chances; it can also give us an empirical basis to work out what kind of sport will make up most of a nation's total medals won.

2.2.4 Distribution of numerical indicators in different sports programs

In order to explore more on the cross-sports differences of the selected numeric indicators, we compare its mean among sports and plot it as bar graphs with uncertainty. which allows for both average levels and variation within the indicator at a glance making this an acceptable basis for those who have either very large or small amounts in these types of discipline.

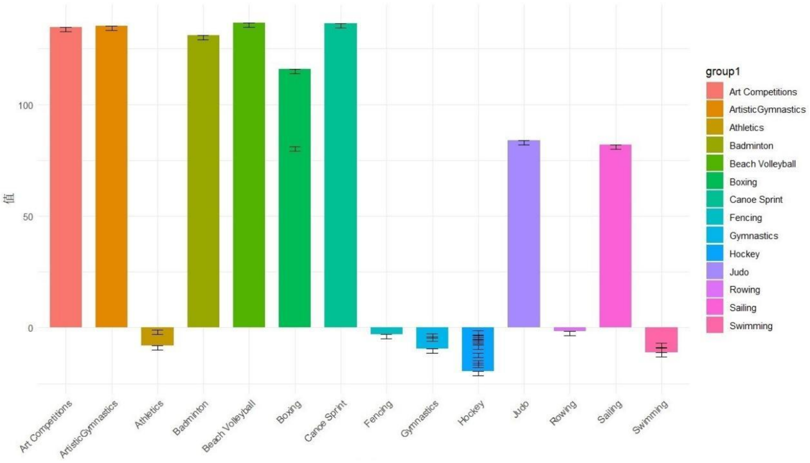

Figure 9 Mean Values of the Numerical Indicator across Sports

As seen in Figure 9, there is great heterogeneity across the disciplines. Art Competitions and artistic gymnastics have high mean values, as do badminton, beach volleyball, canoe sprints; athletics, fencing, gymnastics, hockey, rowing, sailing and especially swimming all fall relatively low on the scale compared to them. These patterns indicate that this numerical indicator isn't distributed evenly across the sports and also points out some discipline - specific differences.

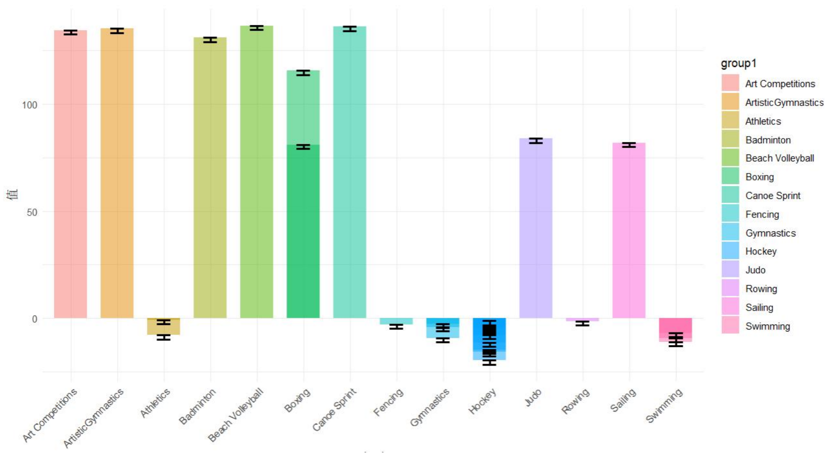

Figure 10 Mean Values with Uncertainty Intervals across Sports

Figure 10 is also a proof of that comparison. Uncertainty of the sports mean The width of the interval implies the instability among indicators in different areas. Few sports showed relatively short intervals suggesting not much scattered observations, on the contrary some have larger ones indicating more scattered observations. Swimming continues to be at the bottom end with only a little spread and hockey has higher uncertainty ranges. On the whole, from Figure 9 and Figure 10 it is jointly verified that this number does not evenly distribute over all sorts of sports. But its mean as well as variability differs per discipline showing we must take cross sport difference into account in further interpretation and modeling.

2.3 Medal Prediction Model

Based on the screening outcomes in section 2.1.2, samples were categorized into 3 categories, i.e., high-performing countries, medium-performing ones, and those with no history of medals: The corresponding model is then used to predict their number of gold medals and overall number of medals won in the 2028 Los Angeles Olympics Games This kind of grouping is due to the differences between the 3 groups when it comes to stability, volatility, data completeness. High-performance nations use a long-term medal share trend , medium-performance nations use time-weighted estimation with shrinkage adjustment , and non-medal nations use a two-stage framework that predicts both the likelihood of winning their first medal and the number of such medals they would win accordingly We have developed a share-smoothing model, a weighted least squares shares regression model, as well as a two-stage model combining logistic regression method and random forest method.

2.3.1 Medal prediction for high-performing countries

For high-performing countries, this study adopts a smoothing prediction model based on linear trend fitting of within-group medal shares. Rather than directly modeling the absolute number of medals, national performance is represented by each country's relative share of the total medals won by all high-performing countries. Let $G_{i,t}$ and $M_{i,t}$ denote the number of gold medals and total medals won by country i in year t, respectively, and let G_t^H and M_t^H denote the total number of gold medals and total medals won by the high-performing country group in the same year. The within-group gold medal share and total medal share are then defined as:

$$s_{i,t}^{(g)}=\frac{G_{i,t}}{G_t^H},\quad s_{i,t}^{(m)}=\frac{M_{i,t}}{M_t^H} \tag{2}$$

Based on these share series, a univariate linear trend is fitted separately for each country. Specifically, the following models are established for gold medal share and total medal share:

$$s_{i,t}^{(G)}=\alpha_i^{(G)}+\beta_i^{(G)}t+\varepsilon_{i,t}^{(G)},\quad s_{i,t}^{(M)}=\alpha_i^{(M)}+\beta_i^{(M)}t+\varepsilon_{i,t}^{(M)} \tag{3}$$

where $\alpha_i^{(\cdot)}$ denotes the intercept term, $\beta_i^{(\cdot)}$ denotes the trend coefficient, and $\varepsilon_{i,t}^{(\cdot)}$ denotes the random disturbance term.

Although these high performing countries are relatively stable, there can still be short term cyclical changes and anomalies in single Olympic Games. So the model does not depend on direct trend extrapolation alone. However, it is also combined with the extrapolated trend using a historical average share via weighted smoothing which allows forecast to contain trend information as well as moderate reverting back toward the long-run mean.

After obtaining the predicted shares for 2028, the model maps them back to the absolute numbers of gold medals and total medals, subject to additional constraints. This mapping can be written as:

$$\widehat{G}_{i,2028}=\hat{s}_{i,2028}^{(G)}\,G_{2028}^{(H)},\qquad \widehat{M}_{i,2028}=\hat{s}_{i,2028}^{(M)}\,M_{2028}^{(H)} \tag{4}$$

2.3.2 Medal prediction for medium-performing countries

In terms of the middle-performing countries, this study uses a share regression model based on weighted least squares principle, which is also added with some conservative extrapolation in regard to those with fewer data. First, countries are screened according to their historical total medals, average medal counts, and peak single-Games performance in order to exclude persistently weak countries and those incorrectly mixed into the high-performing group, thereby forming a relatively homogeneous subsample of medium-performing countries. In this sub-group, every country's share of golds as well as all-medals proportion gets worked out so that forecasting becomes once more about anticipating proportions within the sub-group.

For countries that have participated in at least two Olympic Games, the share series is fitted using a weighted slope estimation method. This method estimates the linear relationship between year and medal share while assigning greater weight to observations closer to 2028. The idea that "observations nearer to 2028 receive larger weights" can be expressed in the weighted least squares framework as:

$$\min_{a_i,b_i}\sum_{t\in T_i} w_t\,(s_{i,t}-(a_i+b_it))^2 \tag{5}$$

For countries that have participated in only one Olympic Games, the historical sample is too limited to support stable estimation of the trend slope. Therefore, weighted regression is not applied; instead, a more cautious conservative extrapolation algorithm is used. At the same time, this study introduces a reliability shrinkage mechanism, under which forecasts are adjusted according to the number of Olympic appearances and cumulative medal achievements of each country: the more complete the historical information, the closer the final forecast is to the raw estimate; the more limited the information, the more conservative the forecast becomes. Finally, the predicted shares of all countries are normalized and allocated within the medal pool of medium-performing countries, while also satisfying upper-bound constraints, tail truncation, and the consistency condition that the total number of medals should not be smaller than the number of gold medals.

2.3.3 Predicting the first medal for historically medal-less countries

For countries that have never won an Olympic medal, this study adopts a two-stage prediction model. The underlying rationale is that the prediction problem for medal-less countries cannot be treated as an ordinary regression problem. Instead, it is first necessary to determine whether such a country is likely to win its first medal in 2028, and then to predict the possible medal count conditional on such a breakthrough.

In the data preparation stage, the sample of historically medal-less countries is first cleaned and then matched with national GDP and population data to construct the candidate feature set. To improve data consistency, country names are standardized, and non-country entities as well as anomalous samples that could not be effectively matched are removed. The final modeling input includes country, year, host effect, GDP, population, and other relevant information.

In the first stage, a logistic regression classifier is used to determine whether a historically medal-less country is likely to achieve its first Olympic medal in 2028. The main explanatory variables include GDP, population size, and host effect, and the logistic function outputs the probability of a first medal for each candidate country. To improve classification stability, median imputation and standardization are performed before fitting the logistic regression model, and the training and test sets are divided using stratified sampling. Because logistic regression provides a clear probabilistic interpretation, it is well suited to the binary classification task of "breakthrough or not."

In terms of classification performance, the first-stage model shows good discriminative ability. On the test set, the classification accuracy is 0.7414, the precision is 0.7222, the recall is 0.7652, the F1-score is 0.6341, and the ROC-AUC reaches 0.8404. These results indicate that the model has strong overall discriminative power, especially given the relatively high ROC-AUC, suggesting that it can effectively distinguish potential breakthrough countries from non-breakthrough countries.

In the second stage, two random forest regression models are employed to predict total medals and gold medals, respectively. GDP, population, historical total medals, historical gold medals, and host effect are used as input variables to predict medal scale in the target year. Random forest regression is appropriate at this stage because it can effectively capture potential nonlinear relationships between explanatory variables and medal counts while maintaining strong robustness.

For the second-stage model evaluation, the regression model for total medals achieves an MAE of 6.5585, an RMSE of 14.5649, and an R^2 of 0.7098; the regression model for gold medals achieves an MAE of 2.1130, an RMSE of 4.0448, and an R^2 of 0.7734. Overall, both regression models exhibit relatively strong goodness of fit.

Finally, the outputs of the two stages are integrated through a threshold rule. If the first-stage probability of winning a first medal falls below a given threshold, both the predicted gold medal count and total medal count for that country are set to zero. If the probability reaches the threshold, the predicted values produced by the two random forest regression models in the second stage are retained, with the additional constraint that the number of gold medals must not exceed the total number of medals. If the breakthrough probability exceeds the threshold but the regression result is still zero, the country is assigned at least one medal to reflect the event-specific nature of a “first medal” achievement.

2.3.4 Summary and interpretation of medal prediction results

Based on the stratified prediction framework described above, this study derives the predicted numbers of gold medals and total medals for each country at the 2028 Los Angeles Olympic Games. The results are summarized through the overall rankings, the largest projected increases and decreases in gold medals, the largest projected increases and decreases in total medals, and the countries most likely to win their first Olympic medal. The corresponding results are presented in Tables 1 through 6.

Table 1 2028 Olympic Games Gold and Overall Medal Standings Predicted Results

No.	Country	2028 Projected Gold Medals	2028 Projected Total Medals
1	United States	60	150
2	China	48	142
3	Great Britain	36	115
4	France	27	102
5	Netherlands	24	70
6	Japan	22	80
...	...	...	...
156	Singapore	0	0
157	Saudi Arabia	0	0
158	Philippines	0	0
159	Puerto Rico	0	0
160	Belarus	0	0

2028 Olympic Games Gold and Overall Medal Standings Predicted Results are shown in Table 1.

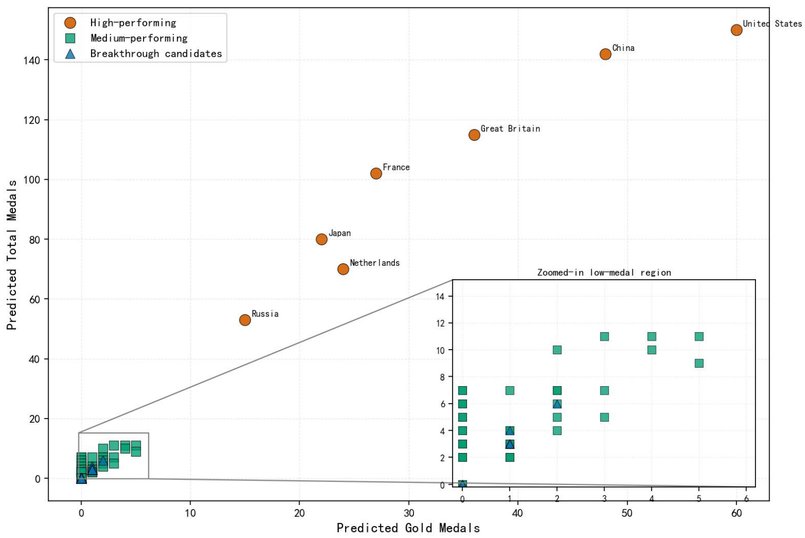

Figure 11 Predicted Gold-Medal Counts Versus Total-Medal Counts across Country Groups

As shown in Table 1 and Figure 11, high-performing countries are expected to remain dominant in the 2028 medal table. The United States, China, Great Britain, France, Japan, and the Netherlands cluster in the upper-right region, indicating strength in both gold medals and total medals. Medium-performing countries lie in the middle range, while breakthrough candidates are concentrated in the low-medal region.

Table 2 Countries with the Largest Projected Increase in the Number of Gold Medals (Top 5)

Country	2024 Actual Gold Medals	Projected Gold Medals	Gold Medal Changes
United States	40	60	+20
China	40	48	+8
Great Britain	14	36	+22
France	16	27	+11
Netherlands	15	24	+9

Countries with the largest projected increase in the number of gold medals is shown in Table 2.

Table 3 Countries with the Largest Projected Reduction in the Number of Gold Medals (Top 5)

Country	2024 Actual Gold Medals	Projected Gold Medals	Gold Medal Changes
Uzbekistan	8	3	-5
Ireland	4	1	-3
Kenya	4	3	-1
Iran	3	2	-1
Georgia	3	2	-1

Countries with the largest projected reduction in the number of gold medals is shown in Table 3.

Table 4 Countries with the Largest Projected Increase in Total Medals (Top 5)

Country	2024 Actual Total Medals	Projected Total Medals	Change in Total Medals
United States	126	150	+24
Great Britain	65	115	+50
France	64	102	+38
China	91	142	+51
Netherlands	34	70	+36

Countries with the largest projected increase in total medals is shown in Table 4.

Table 5 Countries with the Largest Projected Decrease in Total Medals (Top 5)

Country	2024 Actual Total Medals	Projected Total Medals	Change in Total Medals
Iran	12	5	-7
Bulgaria	7	3	-4
Ireland	7	3	-4
Greece	8	5	-3
Turkey	8	5	-3

Countries with the largest projected decrease in total medals is shown in Table 5.

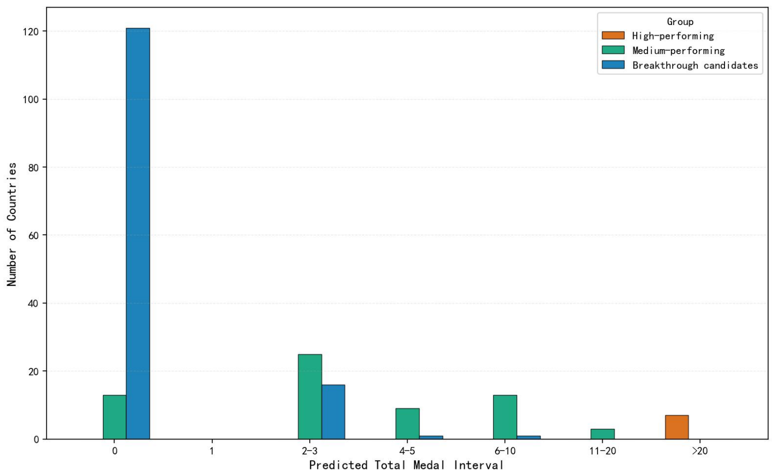

Figure 12 Predicted Gold-Medal Counts Versus Total-Medal Counts across Country Groups

Figure 12 is an obvious demarcation of three performance groups through medal intervals. Top-performing countries congregate in upper echelons; middle-tier ones hover around midscale; and breakthroughs occur on the low end – mostly for those with none, or 0-3 only. This pattern again proves this stratified forecasting model.

Table 6 Countries Predicted to Win their First Medal at the 2028 Olympics (Top 5)

Country	Predicted Gold Medals	Predicted Total Medals	Probability of Winning
Brunei	0	4	0.78
Oman	1	4	0.81
Cambodia	0	1	0.72
Bolivia	0	1	0.59
Bosnia and Herzegovina	0	1	0.56

Countries predicted to win their first medal at the 2028 Olympics (top 5) is shown in Table 6.
Looking at the figure 12, top-performing countries will continue to take most of the medals with the United States, China, Great Britain, France, Japan and Netherlands all predicted to be on top. Due to the host nation advantage, it's expected that the US will have significant gains in both golds and total medals. But for the forecast of average-performing nations, the fluctuation is bigger, its potential improvement also depends greatly on recent competition results, historical accumulations and shrinkage of uncertainty within our model: For historically zero-medal countries, this two-stage model can effectively identify the ones most likely to make a breakthrough (i.e., obtaining at least one medal) by 2028, and provide corresponding predictions regarding their medal winning probability as well as the number of such medals. All things considered, this tiered prediction framework can adapt to stable-type countries, volatile-type countries, breakthrough-type countries separately, so it has a great fit for predicting the medal outcome problem of the 2028 Olympic Games.

3 CONCLUSIONS

By creating a framework that merges multi-level regression analysis with multidimensional feature coupling, this paper completely breaks down the spatiotemporal evolution process of Olympic medals generation. The study finds GDP is the most basic factor of competitiveness for a country at the Olympics while using a hierarchical prediction system handles the fluctuation stability of athletic performance among different nations well Forecast results can effectively identify long-term leading countries like the US and China; and correctly identify potentially breakthrough countries like Brunei, Oman Etc But there were some limits to it in this paper. From the dimensional point of view， it still mainly depends on the macroeconomic indicators nowadays in the existing model， but does not completely take into account the national sports assistance programs implemented by certain countries for some events as well as the competitive standings of individual athletes: In addition to this, the data missing from some places because of geopolitics changes such as in Russia has caused somewhat disruptions in predicting the long trends of the model Future research is intended to add dynamic policy change parameters as well as try out how deep learning tech, which includes LSTM nets, might be combined when creating a world wide strength rating assessment system for competitive sports that can give real time watchfulness and responses.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Kang Huilin, Liu Chunyan. Evolution of the Olympic Medal Tally and the Historical Development of China's International Discourse Power in Sports (2008-2024) // Chinese Society of Sports Science. Hebei Normal University, 2025: 8-9.
[2] Zhang Zhenyu, Liu Chenghao, Xie Yun. Contribution Rate Characteristics of Individual Olympic Triathlon Events and Training Implications // Chinese Society of Sports Science. Tianjin Sport University, 2025: 220.
[3] Ye Zhan, Gao Ping. Comparative Analysis of Olympic Competitiveness Between China and the United States: Statistical Analysis Based on Paris Olympic Results // Hubei Provincial Society of Sports Science. School of Sports Training, Wuhan Sport University, 2024: 594-595.
[4] Ma Qing, Wang Tengyu, Zhao Jing, et al. Research on Olympic Preparation Insights and Strategies for Archery Based on Paris Olympic Results // Sports Information Branch of China Society of Sports Science. Proceedings of the 15th National Sports Information Technology Academic Conference, Tianjin Sport University, 2024: 18.
[5] Fan Xinyu. Analysis of the Preparation Situation for the New Olympic Cycle of the Chinese Swimming Team Based on the Performance of the United States, Australia, and France at the Tokyo and Paris Olympics // Chinese Society of Sports Science, Sports Information Branch. Proceedings of the 15th National Conference on Sports Information Technology, Beijing Sport University, 2024: 24-25.
[6] Lu Haimin. Behind the 4 Billion Views of Chaoxin News' Olympic Short Videos. Media Review, 2024(09): 48-49.

[7] Li Xianjin, Tang Xingxing, Zhang Yang, et al. Comparative Analysis of Competition Results Among Chinese and Foreign Male Artistic Gymnasts During the Paris Olympic Cycle. Sports Science and Technology, 2024, 45(04): 1-4.
[8] Tang Hantao. The “Why Zhejiang” Behind Olympic Achievements. Hangzhou, 2024(15): 68-69.
[9] Hu Guiwei. Chengdu Doctors in Olympic Medical Support Camp: “Seeing Them Achieve Great Results Gives Us a Sense of Accomplishment”. Chengdu Daily, 2024-08-01(005).
[10] Gao Jie, Chen Chen, Li Jiaxin, et al. Dual-Event Characteristics of World-Class Swimmers: An Empirical Study Based on Performance Modeling Across Three Olympic Cycles // Chinese Society of Sports Science. Beijing Sport University, 2023: 224-226.