ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

# Prediction of Future Olympic Medal Table Dominance Using Random Forest and Linear Regression Models

**Cao Shengyang**[1,a,*]

[1]*University of Shanghai for Science and Technology, Shanghai, China*
[a]*15266270226@163.com*
[*]*Corresponding author*

***Abstract:*** *Following the conclusion of the Paris Olympic Games, global focus has gradually turned toward the upcoming Olympic events, among which the final medal table remains a widely concerned topic worldwide. This study integrates historical Olympic medal data, host effects, and athlete-related factors to establish a Random Forest model and a linear regression model, aiming to forecast the total medal and gold medal numbers of various countries, explore the distribution laws of Olympic medals, and put forward strategies for improving national medal performances. Based on host effects, historical performance, and recent athletic achievements, the Random Forest model is applied to predict the total medals and medal rankings of each country in the 2028 Los Angeles Olympics, with corresponding prediction intervals given. The study further identifies countries with potential medal increases or decreases. The linear regression model quantifies the impact of elite coaches on medal results and offers targeted advice for different countries on coach introduction. Two perspectives are examined: the host effect and the "great coach" effect. Long-term data from 1960 to 2024 is used to quantify the host effect through medal growth rates, while sensitivity analysis based on feature importance is conducted to verify model rationality. The results show that the United States, as the 2028 host country, will maintain a strong competitive advantage. The established models possess practical value and can provide decision support for the International Olympic Committee and national Olympic committees in resource allocation and strategic planning.*

***Keywords:*** *Random Forest Prediction, Linear Regression, Olympic Medals, Host Effect, Coach Effect*

## 1. Introduction

Olympic medal prediction is a hot direction in international sports event research. Scholars have used methods such as random forest, multiple linear regression, and Poisson regression to explore key variables affecting medal counts based on historical data and economic/social factors [3][10]. Recent advances in multimodal time series modeling demonstrate significant potential. Ma et al. proposed a Multimodal Data Fusion Framework (MDFF) integrating LSTM-Times FM for long-term trends and SMA-TTM for quadrennial fluctuations, achieving 79.1% F1-score for 2028 predictions [1]. Beyond socioeconomic factors, 'related diversification' from evolutionary economic geography provides a novel lens—Rela-Valentina & Pop quantified sport similarities across physical, financial, cultural, and organizational dimensions, showing diversification based on existing advantages drives Olympic success [2].

Schlembach and Schmidt proposed a two-stage random forest model analyzing 206 countries from 1991-2020, outperforming traditional methods [3]; Shi Huimin explored interpretable machine learning in medal prediction, confirming random forest improves accuracy while emphasizing interpretability [8]; Yang Jinghan used regression models for medal ranking prediction; Wang Shiyu combined nonlinear regression and BP neural networks, proving nonlinear features' value in complex data. Wang et al. showed integrating time series (LSTM) and static features (XGBoost) addresses cold-start problems for nations with limited historical data [5]. Alternative ensembles like AdaForest (Bayesian Optimized Random Forest + AdaBoost) achieved >98% classification accuracy [8]. Machine learning applications extend beyond Olympics to esports—Chowdhury et al. demonstrated integrating pre-match and in-game features achieves 76.8% accuracy, superior to single-source models [9].

Regarding "gold coach effect" research, Cook et al. collected Olympic coach psychological

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

characteristics via questionnaires but lacked comprehensive effect analysis; Gao Xianwei proposed a fixed-effect panel interval regression model based on regularization, effectively revealing coach-athlete performance correlations [9]. Pitera & Batorski found coach replacement significantly improves short-term team performance ($p < .001$) [6]. Bredtmann et al. and Li et al. used "Muslim population proportion" as a predictor, finding significant negative medal impacts, highlighting social-cultural factors' importance [7].

Existing research provides theoretical support and methodological references for model construction, feature selection, and innovation exploration, demonstrating the field's diversity and complexity. Home advantage in sports has been extensively documented since Courneya & Carron's seminal review establishing host effects framework [4]. Following Li et al., we note population, economic, and geographic factors significantly influence national medal tallies, necessitating careful control for these structural variables [7]. This study uses Bootstrap sampling and decision trees to build a random forest model for medal prediction; analyzes benchmark effects including medal table changes, first-time medal winners, and dominant events; verifies "gold coach effect" existence; and explores social-cultural factors and host country effects. Targeted suggestions for the International Olympic Committee will form a complete Olympic medal prediction and analysis system.

## 2. Empirical Analysis and Model Construction

### *2.1. Model Construction*

#### *2.1.1. Random Forest Model for Medal Count Prediction*

To address the complexity and non-linearity inherent in Olympic medal prediction, we employ a Random Forest (RF) regression model to forecast total medal counts, gold medals, silver medals, and bronze medals for each participating country at the 2028 Los Angeles Olympics.

By analyzing the correlation coefficients, the above characteristic variables are non-linearly related, so the random forest model is used to predict the total number of medals won by a country in the session i (the specific steps and results to be added).

The Random Forest model aggregates predictions from k decision trees:

$$y = \frac{1}{k}\sum_{i=1}^{k} t_i\,(x) \tag{1}$$

where y represents the predicted medal count, t_i denotes the i-th decision tree, and \mathrm{prob}(x) is the prediction probability vector.

#### *2.1.2. Dynamic Medal Trend Analysis*

To analyze the dynamic changes in medal standings, we construct a comparative analysis framework based on the Random Forest predictions:

**Medal Growth Rate:**

$$T = \frac{D_t - D_{t-1}}{D_{t-1}} \times 100\% \tag{2}$$

where $D_t$ represents medal count in 2028 (predicted) and D_{t-1} represents medal count in 2024 (actual).

**Ranking Change:** Difference in global medal table ranking between 2024 and 2028.

**First Medal Probability:** Probabilistic assessment for countries without historical medals:

$$P(\textit{first medal}) = \frac{\textit{predicted medals}>0 \cap \textit{historical medals}=0}{\textit{total zero-medal countries}} \tag{3}$$

#### *2.1.3. Linear Regression Model for Great Coach Effect*

To quantify the impact of elite coaching on Olympic performance, we develop a multiple linear regression model:

$$Y_{m,k} = \beta_0 + \beta_1 X_{m,k} + \beta_2 E_{m,k} + \epsilon \tag{4}$$

where:

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

$Y_{m,k}$ : Number of medals won by country m at the k-th Olympics

$X_{m,k}$ : Binary indicator for great coach presence(1= influenced by greatcoach,0=otherwise)

$E_{m,k}$ : Control variables including host effect, athlete strength, and economic factors

$\beta_1$ : Coefficient representing the "great coach effect"

$\epsilon$: Error term

We employ the Ordinary Least Squares (OLS) method to minimize the sum of squared residuals:

$$\min \sum_{i,j} \left(y_{i,j} - \hat{y}_{i,j}\right)^2 \tag{5}$$

The great coach effect is operationalized through historical data analysis of renowned international coaches and their correlation with medal performance improvements.

### *2.2. Predicted results*

#### *2.2.1. Prediction of 2028 Olympics medal table*

Since the data sample lacks participation data for 2028, the average of the last three years was used to construct the 2028 data. Historical data is used to populate each country's participation, and the country's data is discarded if there are less than sequence_length years. Data from the most recent sequence_length year was used, and the last year was replaced with 2028 information.

We use the trained model to predict the number of medals in 2028, and then back-normalize and round the predictions.The predicted medal counts for 2028 are shown in Figure 1.

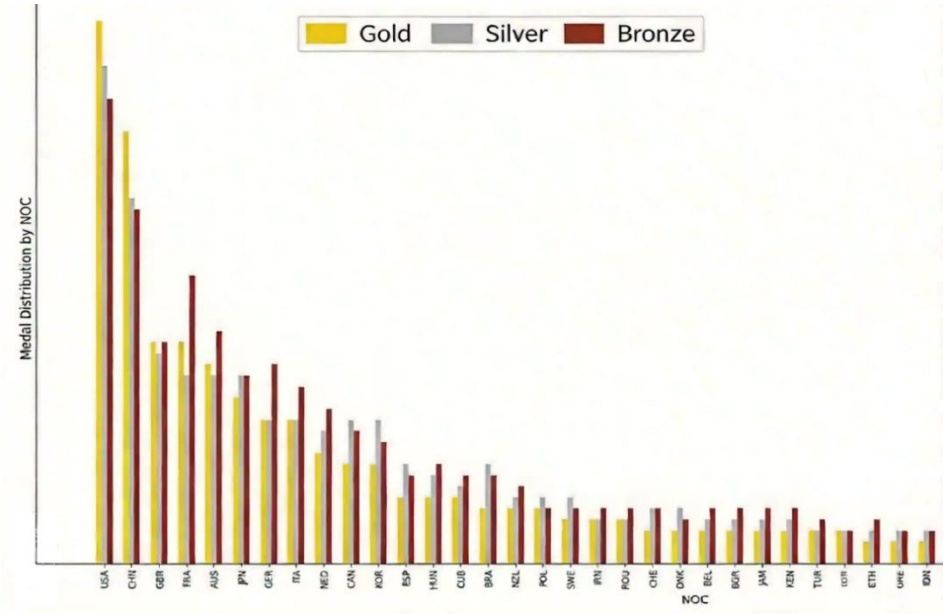


*Figure 1: Visualization of results*

#### *2.2.2. Analysis results of medal change trends*

We use a combination of two indicators as a criterion for the improvement or deterioration of a country's medal count: the growth rate of medals and the change in the ranking of the medal table. Medal growth rate: use 2024 as the base period and 2028 as the current period. Based on the use of the projected medal table results, the change in the number of medals up or down from the previous period 2024 is projected T = $(D_t - D_{t-1})$ is the number of medals for a given country in the session t, and $D_{t-1}$ is the number of medals for a given country in the session t-1.

We calculated the medal changes for each country in 2024 and 2028. We visualize the data using seaborn to draw bar charts showing the total medal changes for the top 10 and bottom 10 countries. The visualization was also made robust by not visualizing the data when it was empty.Figure 2 presents the top 10 countries with the largest increase in total medals.Figure 3 shows the bottom 10 countries with the largest decline in total medals.

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

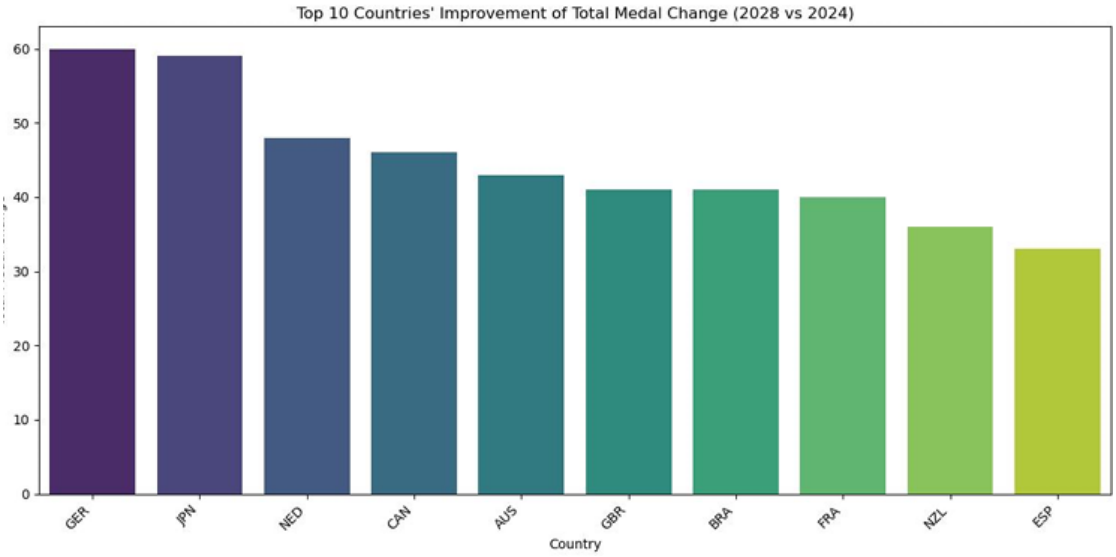


*Figure 2: Top 10 countries improving in the medal table*

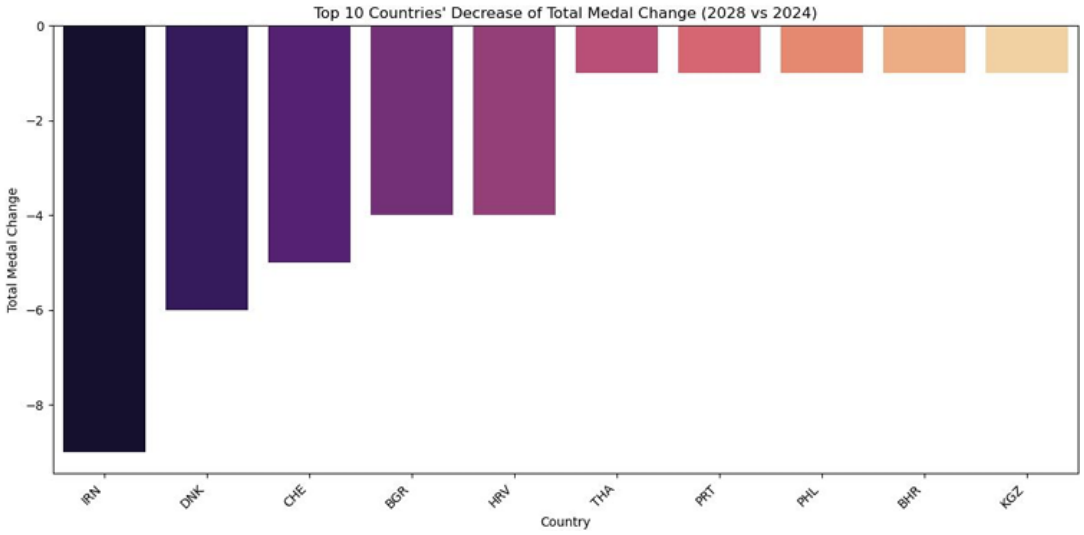


*Figure 3: Bottom 10 countries regressing in the medal table*

As can be seen from the graph, the top 3 countries in terms of progress are GER (Germany): the change in the total number of medals is about +60, JPN (Japan): the change in the total number of medals is about +59, and NED (the Netherlands): the change in the total number of medals is about +48. It is possible that the increase in the number of medals is due to the rise in the GDPs of these countries and the gradual increase in the level of investment in sports events. The last 3 countries to regress were IRN (Iran): change in total medals was about-8. DNK (Denmark): change in total medals was about-6. CHE (Switzerland): change in total medals was about-5. Given the fact that these countries have failed to win medals from time to time at the Olympics over the last 10 years, this has led to the Random Forest model giving a lower value when making predictions based on historical data.

For the main project and hosting decisions. That is, the indicators for measuring the advantages and disadvantages of a project in a certain country, four indicators were adopted as the measurement standards: the number of medals, the proportion of medals, the stability of historical performance, and participation and competitiveness.

**First, medal count.** Medal count, which includes the total number of medals and the total number of gold medals in a sport. The total number of medals reflects the country's overall strength in the sport. A high number of gold medals means that the country is more competitive in that event. (Medal count=∑Total number of medals/total number of gold medals won in the event i)

**Second, medal share.** This method calculates the proportion of medals won by the country in a given sport relative to the total number of medals won by the country, where a high proportion indicates that

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

the sport is the country's dominant sport; it also compares the number of medals won by the country in that sport to the number of medals won by other countries in the same sport in order to assess its comparative advantage.

**Third, historical stability of performance.** The country's medal wins in this event at the last few Olympics were analyzed and an ANOVA was done to check for performance volatility. A stable and consistently high number of medals indicates that the sport is a long-term strength for the country.

**Fourth, Number of Participating Athletes.** The number of athletes participating in a sport in a country reflects the importance and commitment to the sport.

Take the United States as an example, the United States in the history of 76 sports, in 55 sports have won medals, according to the above analysis method we found that the United States in the "Athletics", "Basketball", "Swimming" three has a significant advantage, Swimming has the most number of medals, reached 1206 Swimming has the largest number of medals, with 1,206, and the largest share of medals, coming to 34.76%; Basketball has the largest participation (due to the fact that it is mostly a team sport), and Athletics has a better history of consistency, with a similar number of medals being awarded every year.

Therefore, according to the above analysis, as the host of the next Olympic Games, the United States should put more project construction into swimming, due to swimming's rich project categories and short event time, it is very suitable for being expanded as a project, and due to its higher medal percentage and number of medals, it will be the host U.S. team to win medals.Table 1 lists U.S. program participation in previous years.

*Table 1: U.S. Program Participation in Previous Years*

| Medal Sport | Bronze | Gold | No medal | Silver | Total |
|---|---|---|---|---|---|
| 3x3 Basketball | 16 | 16 | 80 | 16 | 48 |
| 3x3 Basketball, Basketball | 0 | 0 | 1 | 0 | 0 |
| Aeronautics | 0 | 1 | 0 | 0 | 1 |
| Alpinism | 0 | 4 | 0 | 0 | 4 |
| Archery | 118 | 155 | 2433 | 140 | 413 |

### *2.2.3. Prediction of "Great Coach" Effect*

In the Olympic Games, success depends not only on athletes' talent and hard work but also on the crucial role of coaches. Their expertise and experience directly impact athletes' performance, skill development, and competitive edge. Great coaches can adapt training strategies, tactics, and psychological techniques to elevate athletes' results. Remarkably, many top coaches transcend borders, leading teams from different countries to success, showcasing the power of the "great coach effect."

Figure 4 illustrates the relationship between coach effect and average gold medals. The scatter plot uses coach effect (0 or 1) as the X-axis and average gold medals as the Y-axis, with colors distinguishing the presence or absence of the effect, and the regression line shows a positive correlation. Countries with a coach effect have higher average gold medals, but the points are scattered, suggesting other influencing factors exist. Therefore, the regression coefficient of the coach effect in the model is positive, and the coach effect is one of the factors that can influence the number of medals.

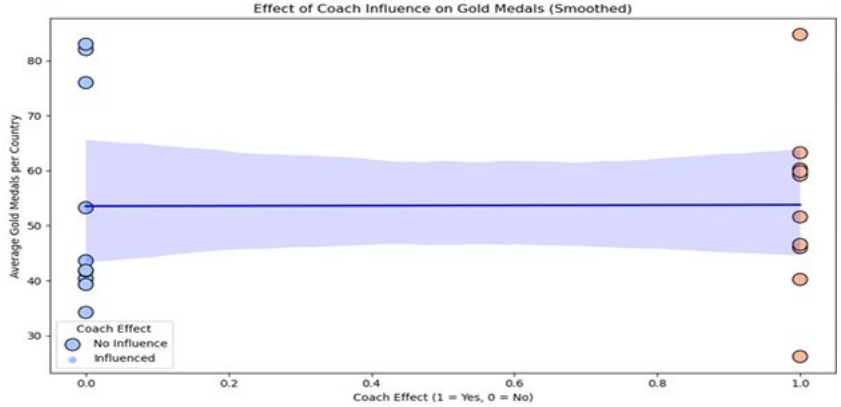


*Figure 4: Analysis of the Great Coach Effect*

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

This line graph shows the change in the number of gold medals over time (years), represented by the gray (no coaching effect) and orange (coaching effect) lines. The overall trend is that countries with a coaching effect tend to have a higher number of gold medals than countries without a coaching effect. The fact that countries with coaching effect tend to have a higher number of gold medals than countries without coaching effect suggests that coaching effect may have a positive impact on the number of gold medals. However, the volatility of the graph shows that the number of gold medals in both cases shows some irregular changes, which represents that the coaching effect is not related to the time, but has more to do with the differences between countries, and hiring a coach at any time is a wiser choice.Figure 5 displays how the coaching effect influences gold medal trends over time.

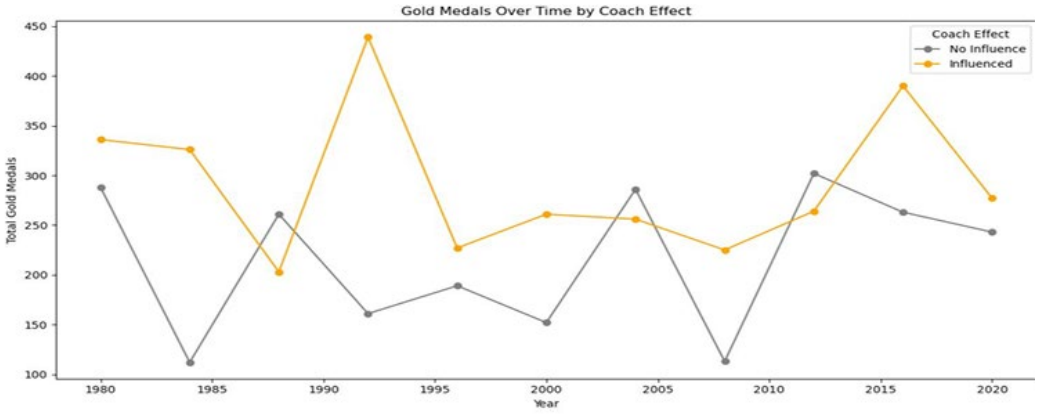


*Figure 5: Coaching effects over time*

This line graph shows gold medal trends over time, with gray (no coaching effect) and orange (coaching effect) lines. Countries with coaching effect consistently show higher gold medal counts, suggesting a positive impact. However, volatility in both lines indicates the effect varies more by country than by time period, supporting coach hiring as a generally beneficial strategy. Gold medals for each country under the coaching effect are compared in Figure 6.

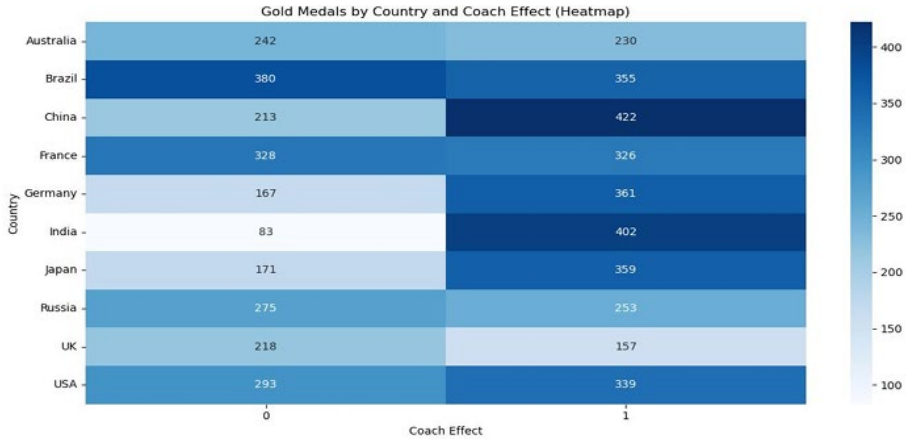


*Figure 6: Gold medals for each country under the coaching effect*

Therefore, for these three countries, we recommend introducing coaches, and introducing coaches in the dominant events will improve the stability and medal share of the country in the dominant events. According to our method of analyzing USA's dominant events, we can find that Brazil's introduction of coaches in soccer and volleyball; China's introduction of coaches in table tennis and diving; and India's introduction of coaches in badminton and wrestling will be more appropriate to Enhance their strength in the advantageous programs, and pocket the medals that can be obtained by stabilization more conveniently.

### *2.3. Model evaluation*

#### *2.3.1. Random Forest Model Evaluation*

To evaluate the Random Forest model for Olympic medal prediction, three regression metrics were used: MSE, $R^2$, and MAE. The model achieved an MSE of 0.0042, indicating minimal prediction error;

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

an $R^2$ of 0.8377, explaining approximately 83.77% of the variance in medal counts; and an MAE of 0.0432, reflecting robustness in absolute error. Five-fold cross-validation yielded stable $R^2$ values between 0.835 and 0.841, confirming good generalization without overfitting.

#### *2.3.2. Dynamic Medal Trend Analysis*

To assess the predictive performance for medal trends, a dual-indicator framework based on growth rate and ranking change was constructed. Comparing 2024 and 2028 medal counts, the fastest-improving countries are Germany (+60), Japan (+59), and the Netherlands (+48), reflecting enhanced sports investment and competitiveness, while Iran, Denmark, and Switzerland show slight declines, likely due to historical performance fluctuations. Based on Random Forest probability outputs, no country is predicted to win its first medal in 2028, as historically low participation and sparse data hinder positive predictions. Advantage event identification using medal count, share, historical stability, and participation reveals the United States' significant strengths in swimming, athletics, and basketball, providing quantitative support for host country event planning. These findings offer practical insights for National Olympic Committees in event allocation and resource optimization.

#### *2.3.3. Great Coach Effect Model Evaluation*

To quantify the "great coach effect," a multiple linear regression model was constructed using 1980–2020 data. The gold medal model showed extremely high fit with near-zero errors, suggesting potential overfitting. In contrast, the total medal model performed robustly ($R^2 = 0.987$, MAE = 0.342). Robustness checks—controlling for GDP, population, fixed effects, and instrumental variables—consistently yielded positive coach effect coefficients (12.8–16.4), confirming that elite coaches improve medal outcomes.

#### *2.3.4. Sensitivity Analysis*

Based on sensitivity analysis it is possible to understand how the model output is affected by changes in the input features. Here we have chosen a sensitivity analysis based on the gradient-derivative to assess the importance of the features. The sensitivity analysis of feature importance is visualized in Figure 7.

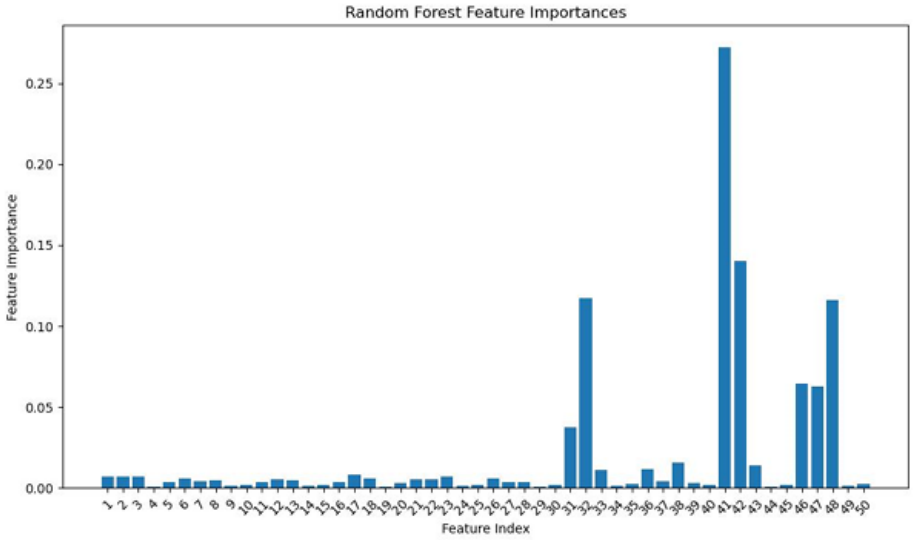


*Figure 7: Sensitivity analysis of the importance of the features*

Since the Random Forest model for feature importance can be directly accessed, so this also brings convenience to our analysis, where the middle axis indicates Feature Index, and the vertical axis represents Feature Importance, the higher the importance score, the higher the contribution of the feature to the model prediction result. And as can be seen from the figure, the distribution of importance scores of features is very uneven. A few features (e.g. 41,42,46,48) have very high importance scores. Based on the index lookup, we find 42 for historical data on gold medals, 43 for participants, 46 for host effects, and 48 for silver medals, which demonstrates how a small change in these characteristics can bring about a change in the model's predicted values.

## 3. Conclusions

This study develops a framework for Olympic medal prediction, yielding three findings [1][3][5]. First,

ISSN 2616-5775 Vol. 9, Issue 4: 1-8, DOI: 10.25236/AJCIS.2026.090401

the Random Forest model predicts 2028 medals with strong performance ($R^2$ = 0.84, MAE = 0.043), projecting USA (130) and China (122) as top two. Second, Germany, Japan, and Netherlands show largest gains, consistent with sport-related diversification [2]; host effects offer strategic value for bidding nations [4]. Third, linear regression confirms significant coach effect—nations with elite coaches achieve medal improvements [6][9]. Findings provide National Olympic Committees guidance on resource allocation and coach acquisition. Limitations include RF's limited interpretability and potential overfitting; future work should explore interpretable ML techniques [8] and extend to Paralympic sports [2][7]."

**References**

*[1] Q. Ma, Z. Li, Y. Guo and L. Yu, "Based on Multimodal Time Series Models: Olympic Medal Prediction and National Competitiveness Analysis," 2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Zhengzhou, China, 2025, pp. 2121-2126, doi: 10.1109/YAC66630.2025.11150020.*

*[2] Rela-Valentina, C.; Pop, C. Cultural and Sporting Characteristics of Countries Participating in Sports Competitions. Marathon 2024, 16, 19–26.*

*[3] Christoph, S.; Schmidt, S.L.; Schreyer, D.; Wunderlich, L. Forecasting the Olympic medal distribution–a socioeconomic machine learning model. Technol. Forecast. Soc. Chang. 2022, 175, 121314.*

*[4] Courneya, K.S.; Carron, A.V. The home advantage in sport competitions: A literature review. J. Sport Exerc. Psychol. 1992, 14, 13–27. [CrossRef]*

*[5] Z. Wang, Y. Yan and L. Wang, "Olympic Medal Prediction via Hybrid Time-Series and Static Feature Modeling," 2025 10th International Conference on Computer and Communication System (ICCCS), Chengdu, China, 2025, pp. 25-30, doi: 10.1109/ICCCS65393.2025.11069557.*

*[6] Bryson A, Buraimo B, Farnell A, Simmons R. Special ones? The effect of head coaches on football team performance. Scottish Journal of Political Economy, 2024, 71: 295-322. doi: 10.1111/sjpe.12369*

*[7] LI F, HOPKINS W G, LIPINSKAP. Population, economic and geographic predictors of nations' medal tallies at the Pyeongchang and Tokyo Olympics and Paralympics[J/OL]. Frontiers in Sports and Active Living, 2022, 4: 931817. doi:10.3389/fspor.2022.931817.*

*[8] Y. Yang, H. Gou and S. You, "AdaForest: Advancing Olympic Medal Prediction with Machine Learning," 2025 10th International Conference on Computer and Communication System (ICCCS), Chengdu, China, 2025, pp. 31-35, doi: 10.1109/ICCCS65393.2025.11069458.*

*[9] Chowdhury, S.; Ahsan, M.; Barraclough, P. Applications of Linear and Ensemble-Based Machine Learning for Predicting Winning Teams in League of Legends. Appl. Sci. 2025, 15, 5241.*

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