Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 5, No. 2, 2025

Medal Prediction for the Olympic Games Based on Grey Prediction Model and Markov Chain

Chuxin Liu

College of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, China

Abstract: This study aims to predict the medal counts for the Olympic Games by employing a combination of the Grey Prediction Model and Markov Chain. The Grey Prediction Model is utilized to forecast the total medal counts for various countries, taking into account multiple covariates such as the number of participating athletes, the number of events participated in, the score of national athlete performance, and the top five sports specialization scores. These covariates are derived from historical Olympic data through meticulous data pre-processing and feature engineering steps, which include handling missing values, standardizing athlete names, and conducting Monte Carlo simulations to estimate participation uncertainties. The Markov Chain model is then applied to predict the likelihood of countries winning their first Olympic medals, based on historical medal outcomes and relevant covariates. The results of the study provide projected rankings for the total and gold medal tables in the 2028 Los Angeles Summer Olympics, as well as estimates for the number of first-time medal-winning countries. Additionally, the study examines the relationship between sports specialization and medal success by analyzing the coefficients of the Grey Prediction Model. The findings highlight the significance of strategic event selection and specialization in achieving Olympic success and offer valuable insights into the competitive landscape of the upcoming Games.

Keywords: Olympic Games; Medal Prediction; Grey Prediction Model; Markov Chain; Sports Specialization.

1. INTRODUCTION

The Olympic Games, as the world's premier sporting event, have always been a focal point for nations to showcase their athletic prowess and competitive spirit. Predicting the medal counts for the Olympic Games is not only of significant interest to sports enthusiasts but also holds substantial implications for sports policymakers, athletes, and coaches. Over the years, numerous studies have attempted to forecast Olympic medal outcomes using various methodologies, ranging from simple statistical models to more complex machine learning algorithms [1, 2]. These studies have highlighted the importance of factors such as the number of participating athletes, the number of events contested, and historical performance data in predicting medal success. However, most existing models either focus on a single aspect of the competition or fail to account for the dynamic nature of international sports performance, which is influenced by multiple interrelated factors [3].

This study aims to address these limitations by developing a comprehensive predictive framework that leverages the strengths of both the Grey Prediction Model and the Markov Chain model. The Grey Prediction Model is particularly effective in dealing with systems characterized by uncertainties, incomplete information, or limited data samples, making it suitable for forecasting medal counts in the Olympic Games [4]. By incorporating multiple covariates, including the number of participating athletes, the number of events participated in, the score of national athlete performance, and the top five sports specialization scores, this model captures the multifaceted nature of Olympic competition. Additionally, the Markov Chain model is employed to predict the likelihood of countries winning their first Olympic medals, providing insights into the potential breakthroughs of emerging nations in the global sports arena [5].

The primary objectives of this study are to: (1) develop a robust predictive model for estimating the total medal counts for various countries in the upcoming Olympic Games; (2) identify the key factors influencing medal outcomes and analyze their interrelationships; and (3) provide projections for the medal table, highlighting the potential shifts in the competitive landscape. The study employs a rigorous data pre-processing and feature engineering approach to ensure the quality and reliability of the input data, followed by the application of the Grey Prediction Model and the Markov Chain model for forecasting [6, 7]. The results of this study offer valuable insights into the potential performance of leading nations and emerging contenders in the 2028 Los Angeles Summer Olympics, emphasizing the importance of strategic event selection and specialization in achieving Olympic success. The findings also highlight the potential for emerging nations to establish themselves as competitive contenders on the global stage, contributing to a more diverse and dynamic Olympic landscape.

2. Data Pre-processing and Feature Engineering

2.1. Data Pre-processing

To ensure the quality and reliability of the data used in our models, we conducted a series of pre-processing steps on the historical Olympic datasets. These datasets encompass comprehensive information on Olympic events and athletes' participation and performance across various Games. Specifically, the data includes detailed records of athletes' names, countries, sports, events participated in, and medal outcomes, among other relevant metrics.

Initially, we addressed the issue of missing values in the dataset. For entries related to S3 (Sailing and Rowing in the 1896 Games, which were cancelled due to adverse weather conditions) and S5 (ice sports such as figure skating and ice hockey, which were part of the Summer Olympics prior to 1924 but later moved to the Winter Olympics), we set the missing values to zero. For other missing values, we inferred and populated the corresponding year and event counts using the available data on athletes' participation. Subsequently, we handled anomalies in the dataset. To standardize the data, we adjusted the names of Chinese athletes in the dataset to ensure the surname consistently appears first. This adjustment was implemented to prevent duplicate entries for the same athlete resulting from inconsistent name formatting.

Finally, to address the uncertainty surrounding the participation of countries in specific events during the 2028 Olympic Games, we employed a Monte Carlo simulation. This simulation leveraged historical data from the past 10 Olympic Games and involved the following steps:

1. For each country, weights were assigned to different sports based on their historical participation frequencies over the past 10 Olympic Games. These weights reflect the likelihood of a country participating in a specific sport in 2028.

2. The simulation was executed 100 times, with each run randomly allocating each country's participation in events according to the assigned weights. This stochastic approach captures the inherent variability and uncertainty in participation patterns.

3. Upon completing the 100 simulation runs, the average participation for each country across all runs was calculated. This average value serves as the estimated participation parameter for 2028.

Through these pre-processing steps, we enhanced the consistency and reliability of the dataset, thereby laying a solid foundation for the subsequent analyses and modelling processes.

2.2. Feature Engineering

Feature engineering is a critical step in constructing a predictive model for estimating a nation's medal tally in the upcoming Olympic Games [6]. By meticulously analysing and processing raw data, this process extracts covariates that significantly influence the target variable (medal count), thereby providing high-quality, relevant input for the model. This section details the extraction of key features from historical Olympic data, including the number of participating athletes, the number of events participated in, the score of national athlete performance, and the top five sports specialization scores. A detailed description of these features is provided below.

1. Number of Athletes $N_i^a(t)$ This represents the total count of athletes from country i in year t. It reflects the scale of a nation's delegation and serves as a key indicator of its potential competitiveness across multiple disciplines.

2. Number of Events $N_i^e(t)$ This indicates the total number of events contested by country i in year t. A higher value suggests broader participation across various sports, which may increase the likelihood of winning medals in diverse categories.

3. Score of National Athlete Performance $S_i(t)$ This is a composite score derived from the average performance ratings of all athletes from country i in year t. It captures the overall competitive strength of a nation's delegation, incorporating individual athlete performance data from recent competitions. The specific calculation of $S_i(t)$ is detailed in Section 3.2.

4. Top 5 Sports Specialization Scores $Q_i^r(t)$ These scores represent the performance strength of country i in year t in its top five most proficient sports, as determined by historical athlete participation and performance data. Each score $Q_i^r(t)$ (where $r = 1,\dots,5$ is calculated based on the total performance score of athletes from country i in a specific sport, derived from the scoring mechanism described in Section 2.3. These scores provide insights into a nation's specialization and dominance in particular sports, which can significantly influence its overall medal tally.

These features collectively capture the scale of a nation's participation, its competitive strength, and its specialization in specific sports, offering multi-dimensional support for the subsequent grey prediction model. Through systematic feature engineering, the predictive accuracy and interpretability of the model are significantly enhanced, laying a robust foundation for forecasting Olympic medal outcomes.

2.3. Athlete Performance Scoring Mechanism

To evaluate the performance of athletes from country t in preparation for the Olympic Games in year t, a weighted scoring mechanism is applied. The process involves two main steps:

1. For each athlete j from country i, their performance score $s_{ij}(t)$ is calculated based on their results from the most recent four Olympic Games (i.e., $t=1$, $t=2$, $t=3$ and $t=4$. The scoring mechanism is defined as follows:

(a) Temporal Weighting Scheme: The results from the most recent four Games are assigned weights α_p (where $p = 1,2,3,4$) as follows:

$$\alpha_1 = 0.4, \quad \alpha_2 = 0.3, \quad \alpha_3 = 0.2, \quad \alpha_4 = 0.1. \tag{1}$$

Here, p = 1 corresponds to the most recent Games (year $t-1$), and $p = 4$ corresponds to the oldest Games (year $t-4$).

(b) Medal-Based Scoring: For each Olympic Games, the athlete's performance is scored based on their medal outcome:

$$m_{ij}(t-p) = \begin{cases} 0.5 & \text{if gold medal,} \\ 0.2 & \text{if silver medal,} \\ 0.1 & \text{if bronze medal,} \\ 0.05 & \text{if no medal,} \\ 0 & \text{if no participation.} \end{cases} \tag{2}$$

(c) Individual Score Calculation: The overall performance score for athlete j in year t is computed as:

$$s_{ij}(t) = \sum_{p=1}^{4} a_p \cdot m_{ij}(t-p). \tag{3}$$

2. The Score of National Athlete Performance $Si(t)$ for country i in year t is obtained by averaging the individual scores of all participating athletes:

$$S_i(t) = \frac{1}{N_i^a(t)} \sum_{j=1}^{N_i^a(t)} s_{ij}(t). \tag{4}$$

where $N_i^a(t)$ defined in the previous subsection.

The scoring mechanism incorporates three key features to ensure a robust and meaningful evaluation of athlete and national performance. First, the temporal weighting scheme emphasizes recent achievements by assigning higher weights to performances in more recent Olympic Games, reflecting the evolving competitive landscape. Second, the medal-based scoring system differentiates between gold, silver, bronze, and non-medal outcomes, capturing the relative significance of each result. Third, national aggregation provides a comprehensive measure of a nation's overall competitiveness by averaging the individual scores of all participating athletes.

Additionally, the scoring mechanism is extended to calculate the Top 5 Sports Specialization Scores $Q_i^r,\ (r = 1,2,3,4,5)$, which quantify a nation's dominance in its most proficient sports based on historical performance data. These scores are derived by aggregating the performance scores of athletes in each sport, weighted by the temporal weighting scheme and medal-based scoring. This ensures a dynamic and accurate representation of a nation's strengths in specific disciplines.

Collectively, these features enhance the accuracy and relevance of the scoring system, enabling a nuanced understanding of both individual and national performance trends over time, as well as specialization patterns across sports.

3. GM Model for Predicting Medal Counts

Grey prediction is an effective methodology for forecasting in systems characterized by uncertainties, incomplete information, or limited data samples. It identifies similarities and disparities in developmental trends among various factors within a system (known as grey relational analysis) and processes raw data through an Accumulated Generating Operation (AGO) to reduce randomness and highlight underlying patterns. Subsequently, a first-order multivariate differential equation model, the GM(1,N) model, is constructed based on the processed data sequence, enabling the prediction of future trends under the influence of multiple interrelated factors. Unlike the GM(1,1) model, the GM(1,N) model accounts for interactions among multiple variables, making it suitable for addressing more complex systemic problems.

Our goal is to predict the total medal count (including gold, silver, and bronze medals) for country i in year t, denoted as $Y_i(t)$. The eight covariates used are described in Section 3.1. The modelling and solution steps for the GM (1,9) model are as follows:

1. Define the original sequence of the target variable (total medal count) as:

$$Y_i^{(0)} = \left(Y_i^{(0)}(1), Y_i^{(0)}(2), \ldots, Y_i^{(0)}(n) \right), \tag{5}$$

where $Y_i^{(0)}$ represents the total medal count for country i in year k. The original sequences of the covariates are:

$$N_i^{a(0)} = \left(N_i^{a(0)}(1), N_i^{a(0)}(2), \ldots, N_i^{a(0)}(n) \right), \tag{6}$$

$$N_i^{e(0)} = \left(N_i^{e(0)}(1), N_i^{e(0)}(2), \ldots, N_i^{e(0)}(n) \right), \tag{7}$$

$$S_i^{(0)} = \left(S_i^{(0)}(1), S_i^{(0)}(2), \ldots, S_i^{(0)}(n) \right). \tag{8}$$

$$Q_i^{r(0)} = \left(Q_i^{r(0)}(1), Q_i^{r(0)}(2), \ldots, Q_i^{r(0)}(n) \right), \tag{9}$$

where $\mathrm{r} = 1, 2, 3, 4, 5$.

2. Apply the Accumulated Generating Operation (AGO) to the original sequences to obtain the accumulated sequences:

$$Y_i^{(1)}(k) = \sum_{j=1}^{k} Y_i^{(0)}(j), \tag{10}$$

$$N_i^{a(1)}(k) = \sum_{j=1}^{k} N_i^{a(0)}(j), \tag{11}$$

$$N_i^{e(1)}(k) = \sum_{j=1}^{k} N_i^{e(0)}(j), \tag{12}$$

$$S_i^{(1)}(k) = \sum_{j=1}^{k} S_i^{(0)}(j), \tag{13}$$

$$Q_i^{r(1)}(k) = \sum_{j=1}^{k} Q_i^{r(0)}(j), \tag{14}$$

where $\mathrm{k} = 1, 2, 3, 4, 5$.

3. Define the GM(1,9) model using the following differential equation:

$$\frac{dY_i^{(1)}}{dt} + aY_i^{(1)} = b_1 N_i^{a(1)} + b_2 N_i^{e(1)} + b_3 S_i^{(1)} + \sum_{r=1}^{5} b_{r+3} Q_i^{r(1)} \tag{15}$$

where a is the development coefficient, and $b_1, \ldots, b_8$ are the influence coefficients of the covariates. The parameter vector $\alpha = [a, b_1, \ldots, b_8]^T$ is estimated using the Least Squares (LS) method. Define the data matrix B and the constant vector $\boldsymbol{Y}$ as:

$$\mathbf{B}=\begin{bmatrix} -Y_i^{(1)}(2) & N_i^{a(1)}(2) & \cdots & Q_i^{5(1)}(2) \\ -Y_i^{(1)}(3) & N_i^{a(1)}(3) & \cdots & Q_i^{5(1)}(3) \\ \vdots & \vdots & \vdots & \\ -Y_i^{(1)}(n) & N_i^{a(1)}(n) & \cdots & Q_i^{5(1)}(n) \end{bmatrix}, \quad \mathbf{Y}=\begin{bmatrix} Y_i^{(0)}(2) \\ Y_i^{(0)}(3) \\ \vdots \\ Y_i^{(0)}(n) \end{bmatrix} \tag{16}$$

The parameter vector $\boldsymbol{\alpha}$ is then estimated as

$$\alpha=(B^T B)^{-1} B^T \mathbf{Y} \tag{17}$$

4. Substitute the estimated parameters into the differential equation and solve for $\hat{Y}_i^{(1)}(k)$. The solution is given by:

$$\hat{Y}_i^{(1)}(k)=\left(Y_i^{(1)}(1)-\frac{b_1}{a}N_i^{a(1)}(k)-\cdots-\frac{b_8}{a}Q_i^{5(1)}(k)\right)e^{-\alpha(k-1)} + \frac{b_1}{a}N_i^{a(1)}(k)+\cdots+\frac{b_8}{a}Q_i^{5(1)}(k). \tag{18}$$

5. Obtain the predicted values of the original sequence by applying the Inverse Accumulated Generating Operation (IAGO) to the accumulated sequence:

$$\hat{Y}_i^{(0)}(k)=\hat{Y}_i^{(1)}(k)-\hat{Y}_i^{(1)}(k-1), \quad k=2,3,\ldots,n \tag{19}$$

This approach provides a robust framework for predicting Olympic medal counts by leveraging the GM (1,9) model, which accounts for the interplay of multiple variables and captures the dynamic nature of international sports performance.

4. Markov Chain Model for Predicting First-Time Medal Winners

Markov chains are probabilistic models that describe sequences of events where the probability of each event depends only on the state of the previous event. This property, known as the Markov property, makes Markov chains particularly suitable for modelling dynamic systems with state transitions, such as the progression of countries from non-medal-winning to medal-winning status in the Olympic Games. By leveraging historical data on medal outcomes and incorporating covariates such as the number of participating athletes, the number of events, and the national athlete performance score, the Markov chain model provides a robust framework for predicting the likelihood of countries winning their first Olympic medal in the upcoming Games [8]. Notably, the Top 5 Sports Specialization Scores $(Q_{i1},\ldots,Q_{i5})$ are excluded from this model, as it focuses on countries without prior medals, which lack sufficient historical data to derive meaningful specialization scores.

To predict the number of countries that will win their first medal in the next Olympic Games, we define the states of the Markov chain as follow:

State 0: A country has never won an Olympic medal.

State 1: A country has won at least one Olympic medal.

The transition probabilities between these states are estimated based on historical data, considering covariates such as the number of athletes $N_i^a(t)$, the number of events $N_i^e(t)$, and the score of national athlete performance $S_i(t)$. The transition matrix P is defined as:

$$P=\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \tag{20}$$

where p_{01} is probability of transitioning from State 0 (no medals) to State 1 (at least one medal), p_{00} is probability of remaining in State 0, p_{10} is probability of transitioning from State 1 to State 0 (assumed to be 0, as countries do not lose their medal-winning status), p_{11} is probability of remaining in State 1

The transition probabilities are estimated using logistic regression, where the covariates influence the likelihood of a country transitioning from State 0 to State 1. The probability $p01$ for country i in year t is given by:

$$p_{01}(i,t)=\frac{1}{1+e^{-(\beta_0+\beta_1 N_i^{a(1)}(t)+\beta_2 N_i^{e(1)}(t)+\beta_3 S_i(1)}} \tag{21}$$

where $\beta0$, $\beta1$, $\beta2$, and $\beta3$ are coefficients estimated from historical data.

Finally, using the estimated transition probabilities, we simulate the Markov chain to predict the number of countries transitioning from State 0 to State 1 in the next Olympic Games. The expected number of first-time medal-winning countries is:

$$\text{Expected Count}=\sum_{i\in\text{State 0}} p_{01}(i,t). \tag{22}$$

5. GM Model for Predicting Medal Counts

Table 1 Top Five Most Proficient Sports by Country (Selected Examples)

Country	Sport 1	Sport 2	Sport 3	Sport 4	Sport 5
AUS	Swimming	Athletics	Hockey	Rowing	Cycling
CHN	Gymnastics	Swimming	Diving	Table Tennis	Athletics
FRA	Fencing	Athletics	Gymnastics	Cycling Swimming	
GBR	Athletics	Rowing	Swimming	Cycling	Sailing
ITA	Fencing	Gymnastics	Athletics	Rowing	Cycling
JPN	Gymnastics	Swimming	Athletics	Judo	Wrestling
KOR	Archery	Handball	Gymnastics	Fencing	Shooting
USA	Athletics	Swimming	Basketball	Rowing	Gymnastics
UZB	Boxing	Wrestling	Judo	Athletics	Swimming

Following the feature engineering process outlined in Section 2.2, we initially identified the five most proficient sports for each country. Due to space constraints, Table 1 presents a subset of countries along with their respective top five sports, as determined by aggregating historical performance data and applying the scoring mechanism described in Section 2.3.

5.1. Projections for the Medal Table

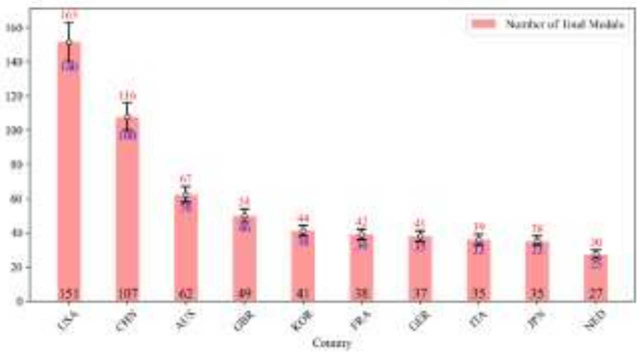

Fig. 1 Projections for the total medal table in LA 2028

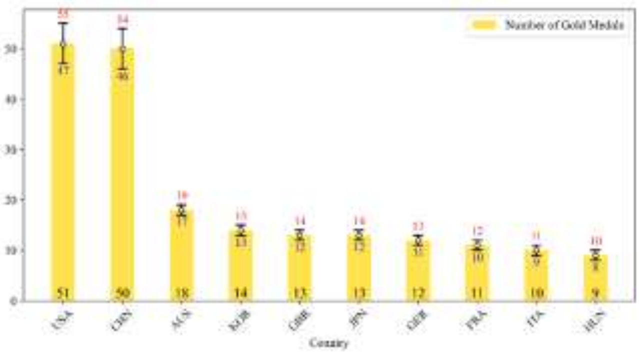

Fig. 2 Projections for the gold medal table in LA 2028

Fig. 1 and Fig. 2 present the projected top 10 rankings for total medals and gold medals, respectively, as forecasted by our model for the 2028 Los Angeles Summer Olympics. These projections offer a glimpse into the potential performance of leading nations and provide valuable insights into the competitive landscape of the upcoming Games.

In the bar charts, the error bars associated with each bar represent the prediction intervals for the total medals or gold medals of each country, as estimated by our model. The red and blue numbers displayed above the error bars indicate the upper and lower bounds of these intervals, respectively. These intervals reflect the uncertainty inherent in our forecasts, outlining the range within which the actual medal counts are expected to fall with a high degree of confidence. The results indicate that the United States, China, and Australia are projected to occupy the top three positions in both medal tables.

Additionally, we examined the relative changes in the total and gold medal counts for the top 17 countries in the medal table, comparing their projected performance in 2028 with their actual results in 2024. These relative changes are summarised in Table 2, which offers a comparative perspective on the evolving competitiveness of these nations over the four-year period.

Table 2 Projected Change in Medal Count Relative to 2024

Country	USA	CHN	AUS	KOR	GBR	JPN	GER	FRA	ITA
Gold Change	11	10	0	1	-1	-7	0	-5	-2
Total Change	25	17	9	9	-15	-10	5	-25	-4
Country	HUN	UZB	NZL	CUB	NLD	CAN	ESP	BRA	Country
Gold Change	3	1	-3	5	-9	-3	1	0	
Total Change	5	1	3	7	-7	-2	1	2	

The United States (USA), China (CHN), and Cuba (CUB) are projected to show the most significant improvements, with increases of 25, 17, and 7 total medals, respectively. In contrast, France (FRA), Great Britain (GBR), and Japan (JPN) are expected to experience declines, with reductions of 25, 15, and 10 total medals, respectively.

5.2. Breakthrough Medal Estimates

Using the Markov Chain model, we estimated the likelihood of countries without prior Olympic medals securing their first medals in the 2028 Olympic Games, with the expected number of medals derived from these probabilities. The part detailed results of these estimates are provided in Table 3.

The projections suggest that nine countries are likely to win their first Olympic medals in the upcoming 2028 Games. Notably, one country (PAR) is expected to achieve two medals, marking a significant milestone in its Olympic history. These predictions underscore the potential for emerging nations to establish themselves as competitive contenders on the global stage.

Table 3 Breakthrough Medal Estimates

Country	No.	Country	No.	Country	No.	Country	No.
AND	0.00	PLE	0.14	SV	0.31	COD	0.54
ARU	0.00	BRU	0.14	MAR	0.32	GUY	0.57
GUM	0.00	GBS	0.15	GAB	0.35	VIE	0.57
HAT	0.00	LBA	0.16	MAD	0.36	LUX	0.58
HON	0.00	MTN	0.16	ZIM	0.36	NCA	0.59

5.3. Relationship Between Sports Specialisation and Medal Success

The coefficients of the Grey Prediction Model GM (1,9) for selected countries are summarised in Table 4, where a represents the development coefficient and b1 to b8 denote the influence coefficients of various covariates. These coefficients provide valuable insights into the relationship between different factors and medal outcomes.

Table 4 Coefficients of the Grey Prediction Model GM(1,9) for Selected Countries

Country	a	b1	b2	b3	b4	b5	b6	b7	b8
AUS	0.82	0.61	0.57	1.00	0.57	0.62	0.20	0.19	0.97
BRA	0.18	0.97	0.75	0.23	0.69	0.71	0.56	0.90	0.88
CAN	0.16	0.64	0.46	0.93	0.08	0.10	0.40	0.82	0.70
CHN	0.62	0.43	0.53	0.14	0.77	0.95	0.40	0.24	0.76
CUB	0.86	0.62	0.71	0.09	0.43	0.38	0.92	0.51	0.62

The results indicate that specific sports and events have a significant influence on medal outcomes for certain countries. For instance, Australia (AUS) shows a high coefficient for b3 (1.00), which corresponds to the Number of Events. This reflects Australia's strong performance in sports such as Swimming, Athletics, and Rowing, where participation across a wide range of events enhances its medal prospects. Similarly, the United States (USA) exhibits a high coefficient for b3 (1.00), in line with its dominance in Athletics, Swimming, and Basketball, where competing in multiple events increases the likelihood of medal success.

China (CHN) demonstrates a high coefficient for b5 (0.95), associated with the Top 5 Sports Specialisation Scores. This highlights China's focused approach in sports such as Gymnastics, Diving, and Table Tennis, where targeted investment and historical success contribute to its medal outcomes. In contrast, Cuba (CUB) exhibits a high coefficient for b6 (0.92), linked to the Score of National Athlete Performance, underscoring the role of individual athlete excellence in sports like Baseball and Boxing.

The host country's selection of events also plays a crucial role. For example, the Netherlands (NED) displays high coefficients for b6 (0.97) and b8 (0.93), corresponding to the Score of National Athlete Performance and the Top 5 Sports Specialisation Scores, respectively. This indicates the significance of sports like Hockey and Swimming, where the Netherlands has a strong historical performance, and hosting events in these disciplines could further enhance medal outcomes.

These findings emphasise the importance of strategic event selection and specialisation in achieving Olympic success. Countries that leverage their strengths in specific sports, such as Australia in Swimming, China in Gymnastics, and Cuba in Baseball, are better positioned to maximise their medal potential.

6. CONCLUSION

This study has developed a robust predictive framework for estimating Olympic medal counts by integrating the Grey Prediction Model and the Markov Chain model. The Grey Prediction Model effectively captures the dynamic nature of international sports performance through multiple covariates, while the Markov Chain model provides insights into potential breakthroughs by emerging nations. The results offer reliable projections for the 2028 Olympic Games, highlighting the importance of strategic event selection and specialization. However, there are areas for improvement. Future work could incorporate real-time data and more granular athlete performance metrics to enhance predictive accuracy. Refining the feature engineering process by including additional variables such as economic indicators and sports infrastructure could also improve model robustness. Exploring advanced machine learning techniques may further refine the predictive capabilities of the models. In summary, this study provides valuable insights into Olympic medal predictions and emphasizes the need for continuous improvement through advanced data integration and model refinement.

REFERENCES

[1] Makiyan S, Rostami M. Economic determinants of success in Olympic Games[J]. Turkish Journal of Sport and Exercise, 2021, 23(1): 33-39.

[2] Orunbayev A. Approaches, behavioral characteristics, principles and methods of work of coaches and managers in sports[J]. American Journal Of Social Sciences And Humanity Research, 2023, 3(11): 133-151.

[3] Goretti L. Athletes' Early Responses to the War Against Ukraine[M]//The Geopolitical Economy of Sport. Routledge,

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

[5] Jablonsky J. Ranking of countries in sporting events using two-stage data envelopment analysis models: a case of Summer Olympic Games 2016[J]. Central European Journal of Operations Research, 2018, 26: 951-966.

[6] Liu S, Yang Y, Forrest J. Grey data analysis[J]. Springer Singapore, Singapore, Doi, 2017, 10(1007): 978-981.

[7] Galeano J, Gómez M Á, Rivas F, et al. Using Markov chains to identify player’s performance in badminton[J]. Chaos, Solitons & Fractals, 2022, 165: 112828.

[8] Galeano J, Gómez M Á, Rivas F, et al. Using Markov chains to identify player’s performance in badminton[J]. Chaos, Solitons & Fractals, 2022, 165: 112828.