The Prediction and Analysis of Medal Counts Based On ALRWI-BPLP

Haitao Li [*], Yuqing Zhuansun

ocean college, Zhejiang University, Zhoushan, China, 316021

* Corresponding Author Email: 3220101495@zju.edu.cn

Abstract. Accurately predicting Olympic medal counts is crucial for evaluating national sporting strength and optimizing resource allocation. Existing prediction methods often fail to fully leverage rich historical information and perform poorly under data-sparse conditions. To address this, this paper proposes a novel hybrid prediction model: For events with sufficient data, a Backpropagation (BP) neural network employing adaptive learning rates (ALR) and Xavier weight initialization (WI) is utilized, incorporating features such as participant numbers from the previous two Games and host nation status for refined prediction. For events with insufficient or no historical data, a combined model primarily based on linear programming (LP), supplemented with Gaussian white noise correction, is adopted. This strategy effectively addresses challenges including zero-value interference in historical data, and data volume disparities, while enhancing prediction robustness. Applied to Olympic data from 1896 to 2024, the model successfully predicted the total medal counts for nations at the 2028 Los Angeles Olympics. The results project: the United States (155 medals), China (82 medals), and France (76 medals) will secure the top three positions on the medal table. Compared to 2024, the medal counts of the United States, France, Australia, and others are projected to increase, while those of China, Great Britain, the Netherlands, and others are expected to decline. Concurrently, forecasting results separately for each sport within a country provides clear and intuitive insights into the nation's degree of reliance on specific sporting events. This enables National Olympic Committees (NOCs) to form expectations regarding potential medal distributions in future Games, thereby facilitating more rational resource allocation to maximize performance outcomes. The findings of this study thus offer a scientific basis and practical reference for Olympic project analysis and strategic decision-making by NOCs regarding resource allocation optimization.

Keywords: Medal Counts Prediction, Backpropagation Neural Network Model, Adaptive Learning Rate, Weight Initialization, Linear Regression.

1. Introduction

As the world’s highest-level international sports event, the Olympic Games have evolved into a top-tier sporting extravaganza encompassing over 200 participating countries and more than 40 major sports disciplines [1-2]. It is not only a concentrated manifestation of the highest levels of competitive sport but also serves as a crucial platform for countries to showcase their comprehensive national strength and resource allocation strategies [3].

Previous research has explored the prediction of Olympic medal performance and the optimization of athletic resource allocation [4]. For instance, certain studies have applied conventional statistical methodologies such as linear regression and time series analysis to predict medal outcomes through historical datasets. In addition, other researchers have investigated the impact of critical factors—including athlete training regimens, sports infrastructure, and governmental funding—on athletic performance [5]. However, these studies often exhibit limitations. A significant number fail to fully leverage the richness of information embedded in historical data. Concurrently, various machine learning models have been employed to predict medal outcomes by identifying patterns and trends within historical datasets [6-8]. ifferent boosting algorithms have demonstrated efficacy in forecasting Olympic medals, underscoring the potential of machine learning in this domain [6]. Furthermore, comparative analyses of machine learning algorithms have yielded valuable insights into their predictive capabilities for Olympic outcomes [7]. Consequently, this study utilizes a comprehensive dataset encompassing all participating athletes and their medal achievements from 1896 to 2024. This study establishes a hybrid model integrating: a Backpropagation Neural Network model with adaptive learning rate and weight initialization, and a composite fitting approach with linear fitting as the primary component, refined by Gaussian white noise correction. This framework capitalizes on both the informational wealth within historical data and the predictive accuracy of machine learning for medal tally forecasting, enabling the prediction of medal counts for nations at the 2028 Olympic Games.

The primary contributions of this study include: 1) Proposing a hybrid model that delivers tailored medal tally forecasts for events with varying data availability; 2) Developing an algorithmic approach to mitigate the impact of zero-value entries in historical datasets (resulting from non-participation of certain nations during specific Olympic cycles); 3) Incorporating an innovative modeling framework that dynamically weights the influence of strong and weak disciplines across nations, accommodating event variations throughout Olympic history.

The structure of this paper is organized as follows: Section I: Introduction delineates the research background, current landscape, and key contributions. Section II: Theoretical Foundations elucidates the linear regression methodology and details the Backpropagation Neural Network algorithm incorporating adaptive learning rate and weight initialization. Section III: Experimental Design and Analysis comprehensively describes the experimental framework and procedural implementation. Section IV: Empirical Findings synthesizes the core research outcomes derived from the experiments. Section V: Conclusion articulates the scholarly significance and future research trajectories.

2. Related Theories

Linear programming (LP) is a mathematical optimization methodology that seeks the optimal solution for a linear objective function subject to a set of linear constraints. Widely applied in resource allocation, cost minimization, and production planning, its canonical formulation comprises three core components: the objective function, decision variables, and constraint equations. Solution algorithms—notably the simplex method and interior-point methods—deliver computational efficiency and intuitive interpretability, making them suitable for large-scale problems. However, due to its inherent linearity assumption, LP exhibits limitations in handling complex nonlinear systems and dynamic constraints.

The backpropagation (BP) neural network algorithm with adaptive learning rate and weight initialization represents a critical improvement over traditional error BP algorithms. It addresses inherent limitations including slow training convergence, susceptibility to local minima, and sensitivity to hyperparameters - particularly learning rate and initial weights. This enhanced algorithm incorporates two core technical innovations: adaptive learning rate mechanisms and scientific weight initialization strategies.

Adaptive learning rate methods (e.g., AdaGrad, RMSProp, Adam) replace fixed learning rates with dynamic, parameter-specific adjustments based on historical gradient information. These methods increase the learning rate in gentle regions with consistent gradient directions to accelerate convergence, while reducing it in steep areas exhibiting gradient oscillations or directional variability to stabilize training.

Scientific weight initialization strategies (e.g., Xavier/Glorot initialization, He initialization) replace basic random initialization. By designing initial weight distributions according to the number of input and output connections per neuron, these methods ensure that during initial training phases: signals maintain appropriate scales during forward propagation, and error gradients avoid exponential vanishing or explosion during backpropagation.

The loss function employed in the BP neural network algorithm in this paper is the Root Mean Square Deviation (RMSD), as shown in Equation (1).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the sample size.

The adaptive learning rate used is shown in Equation (2).

$$lr = \begin{cases} lr, p \le 70 \\ 0.9lr, p > 70 \end{cases} \tag{2}$$

where *lr* is the learning rate (initially 0.001), *p* is the iteration count with an upper limit of 100, and Xavier weight initialization is employed to prevent gradient vanishing. The Xavier initialization uses a standard normal distribution for weight initialization, as shown in Equation (3).

$$\omega = N(0,1) \tag{3}$$

3. Experiments

In this paper, the total medal count of a country is defined as the sum of medals won across different events [9]. The medal count in each event is modeled as the product of the number of participating athletes and the medal-winning rate. Consequently, the framework decomposes into two prediction models: an athlete participation prediction model and a medal-winning rate model.

3.1. Athlete participation prediction model

1) Athlete participation prediction model for events with sufficient data

For traditional sports such as swimming and 100-meter sprint, athlete participation data across nations at the Olympic Games are generally sufficient. Consequently, these events are suitable for modeling using backpropagation neural networks (BPNNs).

Athletes typically enter their competitive golden age between 4–10 years of training. Those within this peak period generally continue to compete in the subsequent Olympic edition. Therefore, participation figures from the previous two Olympic Games provide significant predictive value for estimating athlete counts at the 2028 Games. This justifies implementing a BPNN prediction model with adaptive learning rates and Xavier weight initialization. The output y of each neuron is calculated as shown in Equation (4).

$$y_{output} = linear(\sigma(linear(x_{input}))) \tag{4}$$

where $y_{output} \in R^{1\times1}$, $x_{input} \in R^{4\times1}$, The input vector *x* comprises four variables: *x1*: participation figure from the previous Olympic Games [10], *x2*: participation figure from the Olympic Games two editions prior, *h*: host nation factor [11] (1 for host nations, 0 otherwise), and the bias term (typically set to 1). *linear* denotes the fully connected layer, and $\sigma(x)$ is the sigmoid activation function as shown in Equation (5).

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{5}$$

The flowchart of the BP neural network algorithm for predicting athlete participation is shown in Figure 1.

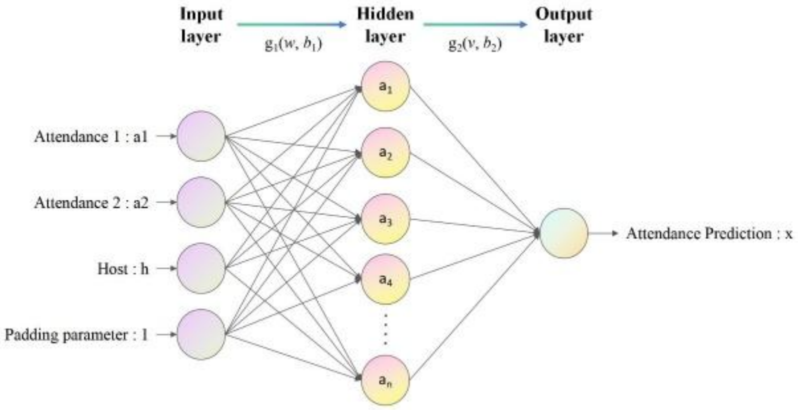

Figure 1. Architecture Diagram of the BP Neural Network for Athlete Participation Prediction

2) Athlete participation prediction model for events with insufficient data

Within the Olympic event roster, certain recently introduced sports – such as rugby sevens – exhibit insufficient historical data. This scarcity leads to significant prediction deviations when fitted with BP neural networks. Consequently, a hybrid model is employed for athlete participation prediction, combining linear regression as the primary fitting mechanism with Gaussian white noise for calibration, as shown in Equation (6).

$$y = ax + b + o(x) \tag{6}$$

where *o(x)* denotes the Gaussian white noise term. The correction value of Gaussian white noise is shown in Equation (7).

$$o(x) = N(0, \sigma^2) \tag{7}$$

where σ is shown in Equation (8).

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - x_i)^2 \tag{8}$$

where n is the sample size.

3.2. Medal-Winning Rate Prediction Model

Athlete caliber varies across different countries, leading to differences in medal-winning rates. To predict the medal-winning rate of various countries in different events, we similarly employed a Backpropagation (BP) Neural Network prediction model based on adaptive learning rate and weight initialization for fitting. The output y of each neuron in the model is shown in Equation (9).

$$y'_{output} = linear^{(2)} x'_{input} \tag{9}$$

where $y'_{output} \in R^{1\times1}$, $x_{input} \in R^{6\times1}$, The input vector x comprises six variables: *A1*: participation figure from the previous Olympic Games, *A2*: participation figure from the Olympic Games two editions prior, *C1*: number of medals won in the previous Olympic Games, *C2*: number of medals won in the two previous Olympic Games, *h*: host nation factor (1 for host nations, 0 otherwise), and the bias term (typically set to 1). *linear* denotes the fully connected layer.

The remaining model architecture is identical to the model described in section 1).

The flowchart of the BP neural network algorithm for predicting medal-winning rates is shown in Figure 2.

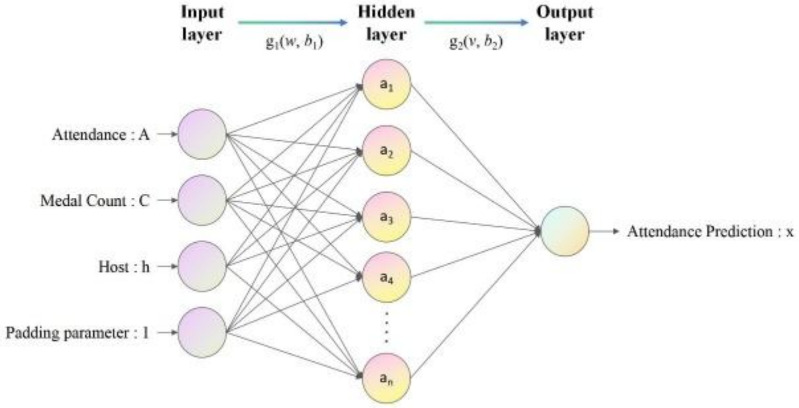

Figure 2. Architecture Diagram of the BP Neural Network for Predicting Medal-Winning Rates

3.3. Prediction Model's Confidence Interval

The width of the prediction interval is determined by the loss function of the enhanced BP neural network. Our model's interval is defined within the 95% confidence interval of the enhanced BP neural network prediction model, as shown in Equation (10).

$$Y \in \left[\hat{Y} - 1.96RMSE, \hat{Y} + 1.96RMSE\right] \tag{10}$$

where n is the sample size.

The total number of medals won by a given country equals the sum of the products of its participant count and medal-winning rate across all events, as shown in Equation (11).

$$\hat{Y} = \sum_{i=1}^{m} y_i y_i' \tag{11}$$

where m represents the total number of Olympic events in which the country participates.

Consequently, the prediction interval for the nation's medal count is shown in Equation (12).

$$Y \in \left[\sum_{i=1}^{m} y_i y_i' - 1.96RMSE, \sum_{i=1}^{m} y_i y_i' + 1.96RMSE\right] \tag{12}$$

The model for predicting a nation's gold medal count is identical in structure to the total medal count model, distinguished solely by the input data: one utilizes gold medal count across all athletes, while the other employs total medal count.

In summary, by constructing the aforementioned objective functions and decision variables, we developed a composite model. This model integrates an enhanced BP neural network—featuring adaptive learning rate and weight initialization—with primary linear fitting, refined by Gaussian white noise correction. Its core objective is to utilize given athlete participation counts and historical medal data to: predict each nation's medal count per event in the upcoming Olympics via the composite model, and subsequently forecast each nation's total medal tally for the next Olympic Games.

4. Results

The medal count for each country in the next Olympic Games is predicted using ALRWI-BPLP. This model primarily employs a BP neural network model with adaptive learning rate and weight initialization, supplemented by linear fitting as the primary fitting method, with Gaussian white noise providing corrections. During the modeling process, for events with sufficient historical data, refined predictions are generated using the BP neural network model with adaptive learning rate and weight initialization. For events with insufficient or no data, the model utilizes linear fitting as the primary method, augmented by Gaussian white noise corrections. This approach overcomes the data limitation constraints of the BP neural network model, thereby yielding more accurate medal predictions.

Taking the US men's swimming team as an example, our medal prediction model yields the prediction curve shown in Figure 3.

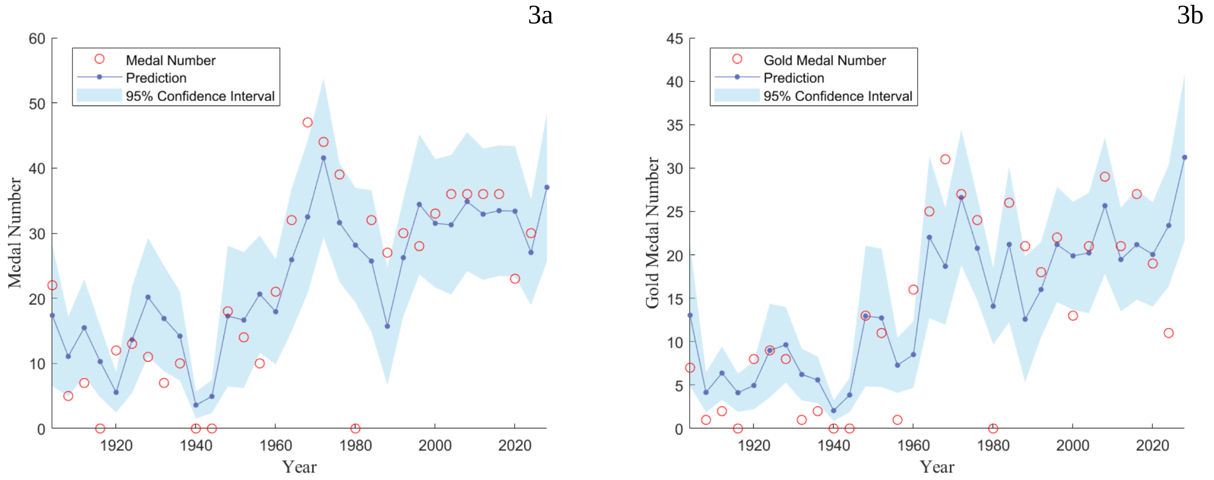

Figure 3. Schematic diagram of predicted (3a) medal count and (3b) gold medal count for U.S. men's swimming.

As evidenced in Figure 3, with the exception of the four years when the United States did not participate in men's swimming events, both medal and gold medal predictions fall within the 95% confidence interval for all other years. This demonstrates the accuracy of our fitting results**.**

Given that our model generates predictions for individual events, it enables medal count analysis across different countries and disciplines. Taking the United States as an example, the medal counts achieved across all events during the 2016 and the 2024 Olympic Games shown in Figure 4.

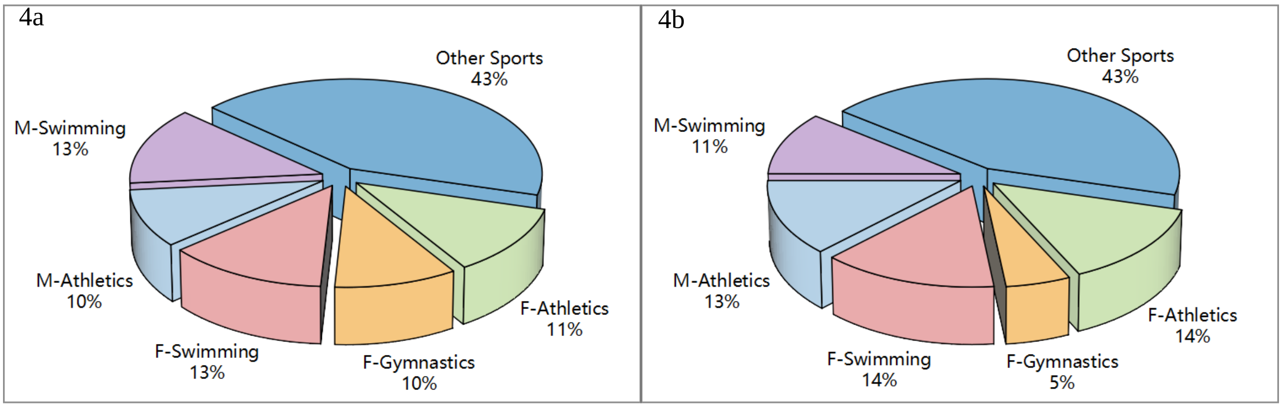

Figure 4. Schematic diagram of the U.S. medal percentage per event in (4a) 2016 and (4b) 2024.

Figure 4 reveals that swimming, athletics, and women's gymnastics collectively account for over 50% of the U.S. medal tally (with highly consistent proportions in 2016 and 2024), establishing a dominant "three-pillar core". This concentration creates critical fragility in the medal ecosystem: our model projects that a 20% reduction in Olympic swimming or athletics events would decrease America's total medal count by 5.0% or 5.3% respectively—exceeding normal fluctuation ranges. This exposes two systemic vulnerabilities: (1) severe resource allocation imbalances where dominant events siphon development resources from other disciplines, and (2) compromised risk resilience where IOC program adjustments (e.g., adding breaking while cutting traditional events) could trigger systemic impacts on the U.S. medal standing.

Diachronic comparison of Figure 4 (2016 vs. 2024) further confirms the persistence of this dependency—the nation has failed to meaningfully diversify its medal sources over the decade. By precisely quantifying event dependencies, our model charts a reform pathway for the USOPC: strategically reallocating resources from the three-pillar core toward emerging disciplines like cycling, fencing, and ball sports (each currently <5% contribution). This resource reallocation would cultivate new medal-growth vectors for sustainable competitiveness. Such predictive capability for data-driven resource optimization constitutes the substantive contribution of this research.

Projected medal counts for all nations at the 2028 Olympic Games are presented in Table 1**.**

Table 1. Table of 2028 Olympic Medals (Forecast) and 2024 Olympic Medals

Country	2024 Total	2028 Total	Prediction interval
United States	126	155	123-187
China	91	82	74-90
France	64	76	60-91
Australia	53	61	52-70
Great Britain	65	56	43-69
Japan	45	56	42-70
Italy	40	46	36-56
South Korea	32	37	31-44
Brazil	20	32	26-37
Netherlands	34	31	25-38

As shown in Table 1, projected increases in medal counts for the United States, France, and Australia at the 2028 Olympics compared to 2024. As host nation, the United States is expected to surge from 126 to 155 medals (a 23% increase), consolidating its top ranking—consistent with historical host-nation advantages derived from event benefits and resource allocation. France (+12 medals) and Australia (+8 medals) also show significant gains, with France overtaking Britain to rank third at 76 medals, reflecting sustained improvement in elite sport development.

Conversely, China (82 medals), Britain (56 medals), and the Netherlands (31 medals) exhibit declines. Despite a 9-medal reduction, China maintains second position, while Britain's 13.8% decrease drops it to fifth—suggesting challenges from emerging nations or internal program restructuring. Nevertheless, traditional powers (US, China, France, UK) retain top-five positions.

Japan (+11 medals) joins Australia and Britain in a tightly contested second-tier cluster (56-61 medal range). Model confidence intervals reveal differential robustness: China exhibits the narrowest range (74-90 medals, span=16), indicating high-fitting historical data from dominant disciplines, whereas Britain's wide span (43-69 medals, span=26) suggests greater uncertainty, potentially due to diversified medal sources or variable new-event participation. The efficacy of our data-driven optimization strategy is validated by accurate predictions for mid-tier nations like Brazil (+12) and Korea (+5).

Cross-referencing Figure 4, America's medal concentration in swimming and athletics (~50% share) indicates host-nation resource amplification in core disciplines. China's adjustment may be attributed to intensified competition or strategic recalibration in key strengths. Our model's innovative discipline-specific forecasting provides strategic insights—for instance, indicating that 20% reductions in US swimming or athletics programs would decrease total medals by >5%, underscoring optimization imperatives.

Projected gold medal counts for all nations at the 2028 Olympic Games are presented in Table 2.

Table 2. Table of 2028 Olympic Gold Medals (Forecast) and 2024 Olympic Medals

Country	2024 Gold	2028 Gold	Prediction interval
United States	40	68	46-69
China	40	40	36-43
France	16	33	25-40
Italy	12	20	16-24
Australia	18	19	16-23
Great Britain	14	17	14-21
Japan	20	17	13-22
South Korea	13	12	9-14
Canada	9	8	6-10
Germany	12	8	5-10

As shown in Table 2, compared to the 2024 Olympics, the gold medal counts of countries such as the United States, France, the United Kingdom, and Italy are projected to increase to varying degrees at the 2028 Olympics. The US gold medal tally is forecast to surge from 40 to 68. Notably, the lower bound of its 95% confidence interval (46-69) still exceeds its 2024 level, highlighting the dual impact of the host nation advantage and dominance in core sports (swimming and track & field). France demonstrates a leap forward (from 16 to 33), potentially driven by the motivational effect of hosting the previous Games on its national athletes. Steady improvements by Italy (from 12 to 20) and the UK (from 14 to 17) further consolidate the competitiveness of European nations within the second tier of the gold medal table.

In contrast, significant declines are projected for Japan (from 20 to 17) and Germany (from 12 to 8). China's gold medal count is expected to remain stable at 40 (interval: 36-43), while Australia (18 to 19), South Korea (13 to 12), and Canada (9 to 8) show minimal change. Notably, despite this divergence in national performances, the three major sporting powers – the United States, China, and France – maintain their positions within the top five of the gold medal standings. China's exceptionally narrow prediction interval (span of 7 medals) reflects the enduring dominance in its traditional strongholds, such as table tennis and diving. However, the stagnation in its total count suggests insufficient breakthroughs in emerging sports.

Japan's projected decrease (-3 medals) may be linked to the dissipation of the 'post-host effect'. Germany's decline (-4 medals) reveals a talent gap risk. The near-stagnant gold medal counts for Australia, South Korea, and Canada underscore systemic bottlenecks in their elite athlete development pipelines. Model predictions for non-traditional powers further illuminate the nature of the challenge: the upper prediction bounds for Canada and Germany (10 medals each) remain below their 2024 levels, indicating structural pressures requiring adjustment within their high-performance sports systems.

The characteristics of the gold medal prediction confidence intervals align with those for total medals: China exhibits the narrowest range (span of 7 medals), while the US has the widest (span of 23 medals). This again confirms the greater predictive robustness for nations with abundant historical data. Combined with Figure 4, the roots of these shifts can be analyzed: The US's dramatic gold medal increase is heavily reliant on swimming and track & field (accounting for nearly 50% of its medals). A reduction in these events would disproportionately impact its gold medal count compared to other nations. France's projected gains may be concentrated in its traditional strengths like fencing and cycling, necessitating vigilance against the risk of over-specialization. These results provide a clear warning to National Olympic Committees: optimizing the portfolio of high-value events and mitigating single-sport dependency will be a core strategic priority for the next quadrennial

5. Conclusions

Leveraging BP neural networks and linear programming algorithms, this study proposes a novel Olympic medal prediction model to forecast medal count for selected leading sporting nations at the 2028 Olympic Games. The results demonstrate that employing distinct fitting algorithms tailored to events with varying data volumes significantly enhances both the accuracy and stability of the predictions. Innovatively, the model incorporates both participant numbers and medal count from the preceding two Olympic Games as initial prediction conditions and conducts event-specific forecasting. This approach ensures that both the strengths and weaknesses of different nations across various disciplines are accurately reflected in the outcomes. Concurrently, forecasting results separately for each sport within a country provides clear and intuitive insights into the nation's degree of reliance on specific sporting events. Thereby enabling National Olympic Committees (NOCs) to form expectations regarding potential medal distributions in future Games, facilitating more rational resource allocation to maximize performance outcomes. The findings of this study offer a scientific basis and practical reference for Olympic project analysis and decision-making by NOCs.

Currently, the model cannot predict the performance of nations that have never previously won an Olympic medal. To address this limitation in future research and model refinement, the inclusion of additional factors potentially influencing Olympic medal distribution should be considered. These factors may include national investment levels in sports, the size of the athlete pool, host nation effects, and cultural preferences toward specific sports. Integrating these variables would enhance the model's generalizability and scalability, enabling more comprehensive predictions of potential performances by a wider range of nations in future competitions.

References

[1] Gould D, Guinan D, Greenleaf C, et al. Factors affecting Olympic performance: Perceptions of athletes and coaches from more and less successful teams[J]. The sport psychologist, 1999, 13(4): 371-394.

[2] Malfas M, Theodoraki E, Houlihan B. Impacts of the Olympic Games as mega-events[C]//Proceedings of the Institution of Civil Engineers-Municipal Engineer. Thomas Telford Ltd, 2004, 157(3): 209-220.

[3] Zhang Z, Ma T, Yao Y, et al. Predicting Olympic Medal Performance for 2028: Machine Learning Models and the Impact of Host and Coaching Effects[J]. Applied Sciences, 2025, 15(14): 7793.

[4] Deepak V, Anguraj D K, Mantha S S. An efficient recommendation system for athletic performance optimization by enriched grey wolf optimization[J]. Personal and Ubiquitous Computing, 2023, 27(3): 1015-1026.

[5] Thompson F, Rongen F, Cowburn I, et al. The impacts of sports schools on holistic athlete development: a mixed methods systematic review[J]. Sports medicine, 2022, 52(8): 1879-1917.

[6] Sagala N T M, Ibrahim M A. A Comparative Study of Different Boosting Algorithms for Predicting Olympic Medal[C]//2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED). IEEE, 2022: 1-4.

[7] Badoni P, Choudhary P, Rudesh C P, et al. Predicting Medal Counts in Olympics Using Machine Learning Algorithms: A Comparative Analysis[C]//2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech). IEEE, 2023: 116-121.

[8] Sayeed R, Hassan M T, Rahman M N, et al. Machine Learning Models for Predicting Olympic Medal Outcomes[C]//2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI). IEEE, 2025, 3: 1-6.

[9] Bernard A B, Busse M R. Who wins the Olympic games: Economic development and medal totals[M]. National Bureau of Economic Research, 2000.

[10] Tian Hui, He Yiman, Wang Min, etc Prediction of Chinese athlete medals and competition strategies for the 2022 Beijing Winter Olympics: based on the analysis of the home advantage effect of the Olympics [J]. Sports Science, 2021, 41 (2): 3-13, 22.

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