Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3051

OLYMPIC MEDAL PREDICTION AND ANALYSIS BASED ON LSTM AND TOPSIS MODELS

DingShu Yan
College of life sciences, Shanxi Agricultural University, Jinzhong 030801, Shanxi, China.
Corresponding Email: yds2004517@163.com

Abstract: In the context of the increasingly fierce global sports competition, accurately predicting Olympic medal outcomes and optimizing the allocation of sports resources have become crucial concerns for national sports committees and related organizations. This study tackles these challenges through the use of advanced modeling techniques. A Long Short-Term Memory (LSTM) model was constructed using historical data from the Summer Olympics (1896-2024), encompassing medal counts, participating events, and national indicators such as population and GDP. The model takes into account the time-dependence of historical performance, advantages in sports infrastructure, and the benefits of being the host country. The results predict that the United States, China, and France will demonstrate strong medal competitiveness at the 2028 Los Angeles Olympics, with potential breakthroughs from emerging nations. Moreover, a decision tree model was employed to examine the influence of "great coaches" on medal results. By examining coach mobility, athlete performance data, and changes in medal counts, the study revealed that transnational coach mobility significantly influences medal distribution. Notable coaches like Lang Ping and Bela Karolyi have enhanced the competitiveness of volleyball and gymnastics, respectively. The findings suggest that recruiting top-tier coaches can increase medal counts and elevate international sports performance. This research provides valuable strategies for optimizing sports resource allocation and enhancing global competitiveness.
Keywords: Olympic medal table; LSTM prediction model; Decision tree model; Great coach effect; Sport resource allocation

1 INTRODUCTION

The 2024 Paris Olympics have highlighted the significance of medal standings, which largely reflect the sporting strength of each country [1]. The United States emerged as the top medal winner, with both China and the U.S. sharing the first place in the total number of gold medals [2]. Some countries achieved their best-ever performances and won their first-ever medals [3]. However, over 60 countries did not win any medals, indicating an extremely unbalanced distribution of medals [4]. Traditionally, predictions were based on the form and strength of individual athletes [5]. Yet, historical data also holds great value and should not be underestimated [6]. Factors such as adjustments in event settings and the coaching level of the coaching team also affect medal distribution. This study mines key information from historical data, constructs a medal distribution prediction model based on it, and considers various factors that interfere with medal distribution to make a more comprehensive and objective prediction of each country's medals [7]. It is hoped that the model can provide targeted suggestions for national Olympic committees, enabling them to plan their sports investment more scientifically and allocate resources rationally, thereby effectively improving the competitive level of their athletes [8].
Previous research has explored the prediction of Olympic medal outcomes and the optimization of sports resource allocation [9]. For example, some research has applied conventional statistical methods like linear regression and time series analysis to predict medal counts using historical data. Others have focused on the impact of specific factors such as athlete training methods, sports facilities, and government funding on sports performance [10]. However, these studies often have limitations. Many of them fail to fully utilize the rich information contained in historical data, such as the time-dependence of historical performance and the advantages of sports infrastructure. Additionally, the impact of transnational coach mobility on medal distribution has not been adequately considered.
This paper overcomes these limitations by utilizing advanced modeling approaches, including long short-term memory (LSTM) models and decision tree models. The LSTM model takes into account the time-dependence of historical performance, sports infrastructure advantages, and the benefits of being the host country, providing more accurate predictions of medal outcomes [11]. The decision tree model analyzes the impact of "great coaches" on medal outcomes, revealing the significant influence of transnational coach mobility on medal distribution [12]. Our findings suggest that recruiting top-tier coaches can increase medal counts and elevate international sports performance, offering valuable strategies for optimizing sports resource allocation and enhancing global competitiveness.
The data in this article comes from https://olympics.com/en/paris-2024/medals.

2 MODEL

2.1 LSTM

The Long Short-Term Memory (LSTM) network is a specialized variant of Recurrent Neural Networks (RNNs) aimed at tackling the challenge of long-term dependencies in sequential data. Standard RNNs often fail to capture long-term dependencies due to problems such as vanishing and exploding gradients. To overcome these limitations, LSTM incorporates a gating mechanism that regulates the flow of information via memory cells, input gates, forget gates, and output gates. This structure enables LSTMs to retain information over extended periods and accurately model intricate temporal relationships.

The LSTM model functions via a set of gates and cell states that manage the flow of information. Its core components consist of:

Forgetting Gate: The forgetting gate decides which information to discard from the cell state. It employs a sigmoid activation function to generate values ranging from 0 to 1. A value of 0 means the information should be entirely forgotten, while a value of 1 signifies that it should be completely retained.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

where f_t is the forgetting gate output, W_f is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, and b_f is the bias vector.

Input Gate: The input gate decides what new information to store in the cell state. It is composed of two components: a sigmoid layer and a tanh layer. The sigmoid layer regulates the update process, while the tanh layer generates a new candidate value to be incorporated into the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\widetilde{C}_t = \tanh(W_C [h_{t-1}, x_t] + b_C) \tag{3}$$

where i_t is the input gate output, $\widetilde{C}_t$ is the candidate cell state, W_i and Wc are weight matrices, and b_i and b_c are bias vectors.

Cell State Update: The cell state is refreshed by integrating the previous cell state with the outputs from the forgetting gate, the input gate, and the candidate cell state.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{4}$$

where C_t is the current cell state.

Output Gate: The output gate determines what information should be output based on the current cell state. It uses a sigmoid gate to control the output and a tanh function to scale the output.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where o_t is the output gate output, h_t is the current hidden state, W_o is the weight matrix, and b_o is the bias vector.

2.2 Decision Tree Model

A decision tree is a non-parametric, supervised learning algorithm applicable for both classification and regression. It operates by recursively splitting the dataset into subsets according to the values of input features, forming a tree-like structure of decisions. In this structure, each internal node corresponds to a test on a feature, each branch indicates the result of the test, and each leaf node signifies a class label or a predicted value.

The decision tree model functions by iteratively dividing the data into subsets according to the values of input features. The essential steps are as follows:

Feature Selection: The model identifies the most effective feature for partitioning the data, using metrics like Gini impurity for classification tasks or mean squared error (MSE) for regression tasks.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the number of samples.

Data Partitioning: The dataset is divided into smaller subsets according to the chosen feature and a specified threshold. This procedure is recursively applied to each subset until a termination criterion is satisfied, such as reaching the maximum tree depth or achieving a minimum number of samples per node.

Tree Construction: The tree is built by repeatedly performing the data-splitting process. Each internal node corresponds to a decision based on a specific feature, while each leaf node indicates a prediction value.

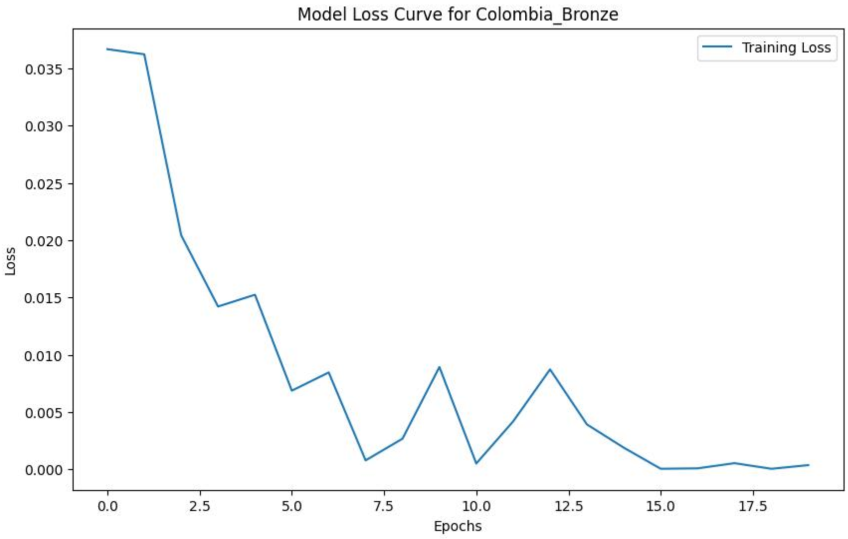
Prediction: For a new input sample, the model navigates from the root node to a leaf node, evaluating the input feature values at each internal node to determine the path. The prediction is the value associated with the final leaf node.

3 RESULTS AND ANALYSIS

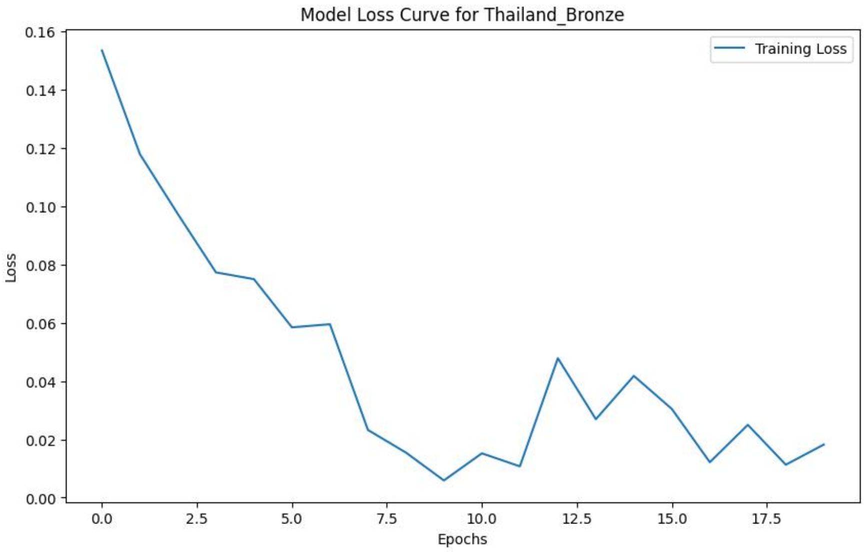
3.1 Results and Analysis of the LSTM Model

The LSTM model was developed using historical medal data from the Summer Olympics (1896-2024), incorporating features like the number of medals won, participation in events, national population, and GDP. It was designed to forecast the medal count for each country in the 2028 Los Angeles Olympics. During training, the data was standardized and normalized using the MinMaxScaler to scale the values between 0 and 1. The model was trained over 20 epochs with a batch size of 32.

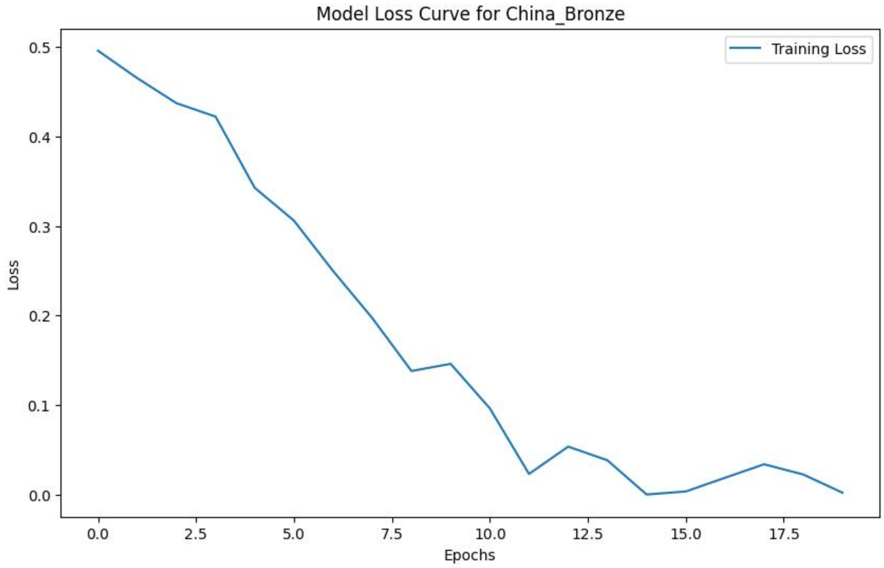
The training loss curve (Figure 1) illustrates a steady decline in the loss value as the number of training iterations increased, suggesting that the model progressively converged and captured the underlying patterns in the data.



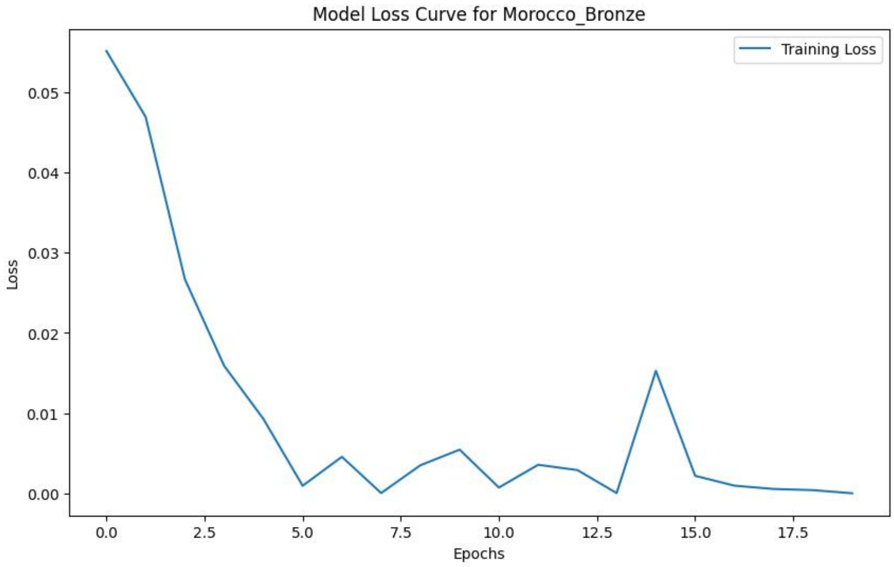
(a)Model Loss Curve for Colombia_Bronze



(b)Model Loss Curve for Thailand_Bronze



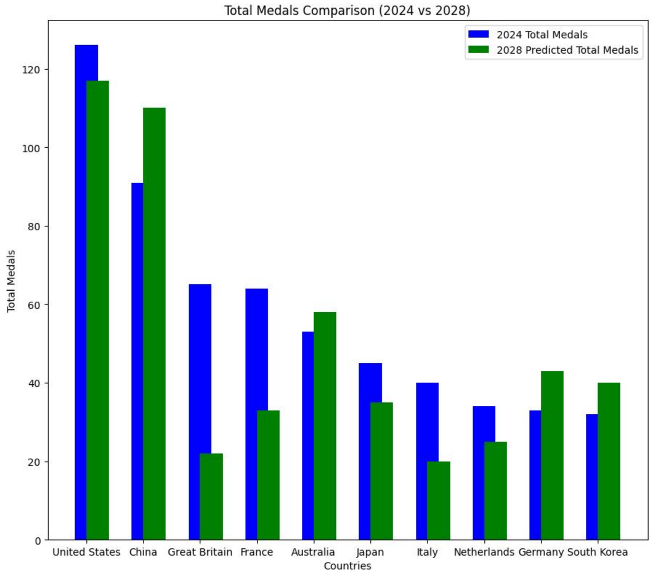
(c)Model Loss Curve for China_Bronze



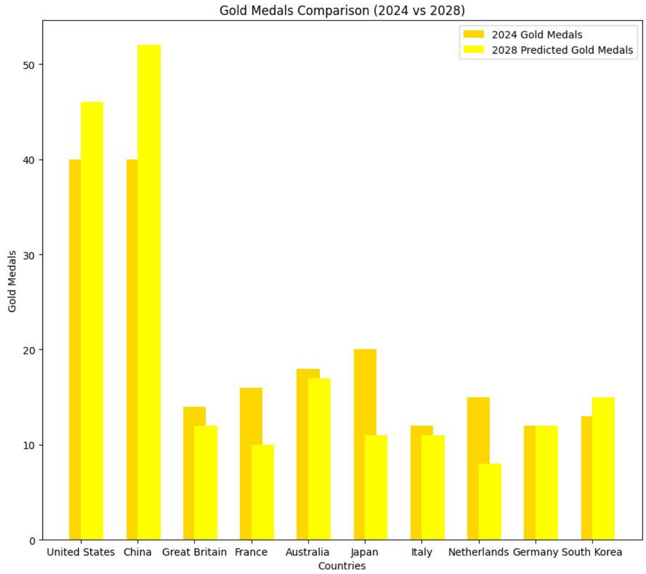
(d)Model Loss Curve for Morocco_Bronze

Figure 1 Training Loss Curve of the LSTM Model

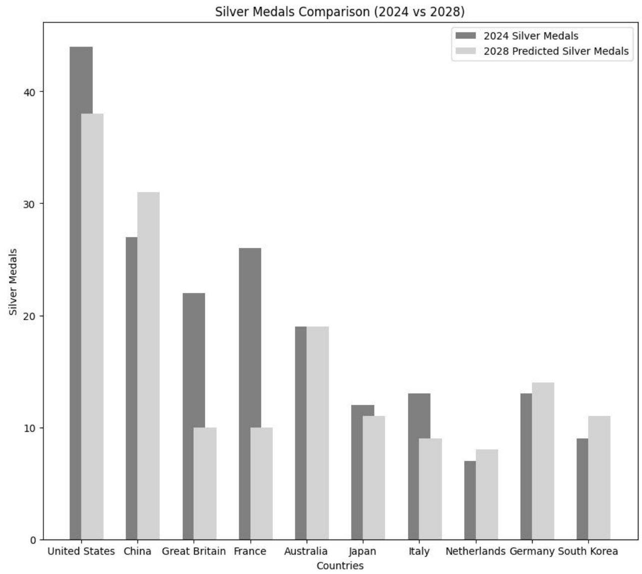
The model's forecasts for the total number of medals, as well as gold, silver, and bronze medals for each country in the 2028 Olympics, were compared with the actual medal counts from the 2024 Olympics. The findings (Figure 2) indicate that the United States and China are anticipated to continue leading the medal standings, with the United States potentially experiencing a modest decline in overall medals. Other nations, including Great Britain, France, and Australia, are projected to maintain relatively stable medal counts.



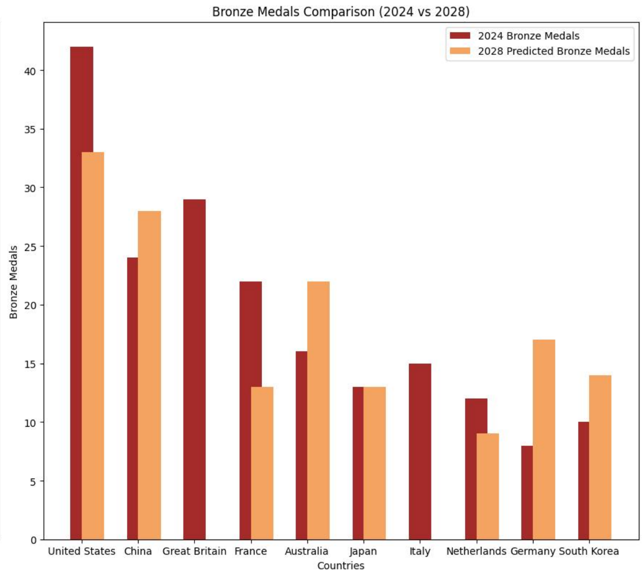
(a)Total Medals Comparison(2024 vs 2028)



(b)Gold Medals Comparison(2024 vs 2028)



(c)Silver Medals Comparison(2024 vs 2028)



(d)Bronze Medals Comparison(2024 vs 2028)

Figure 2 Predicted vs. Actual Medal Counts for Top Countries (2024 vs. 2028)

The model also predicted the medal counts for countries that did not win any medals in the 2024 Olympics. The top ten countries most likely to win medals in the 2028 Olympics are shown in Figure 3, with Finland being the most likely to achieve a breakthrough.

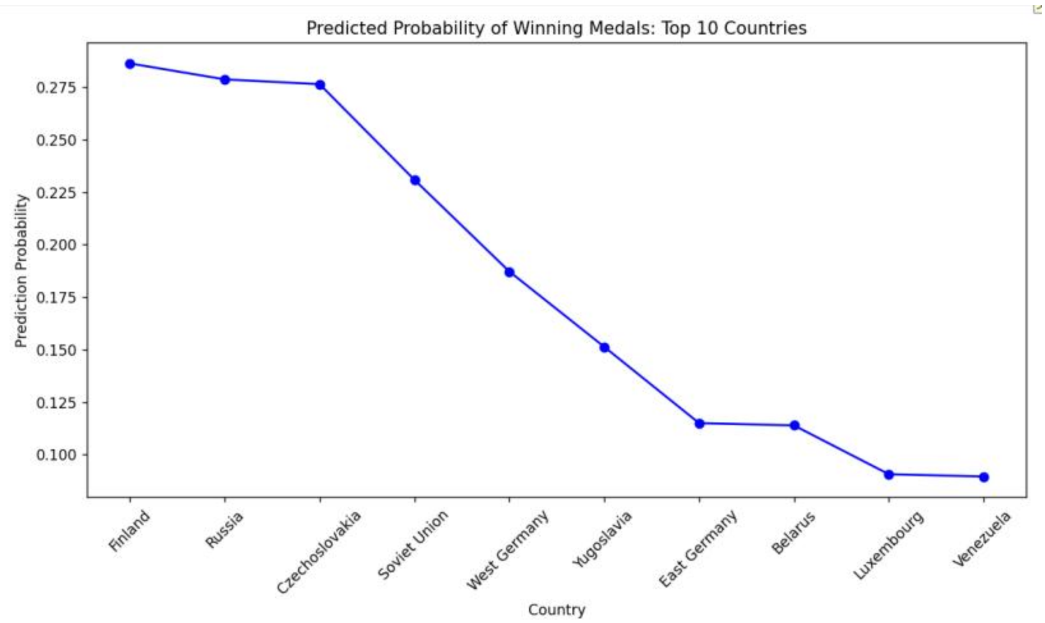


Figure 3 Top Ten Countries Most Likely to Win Medals in 2028

The LSTM model successfully identified the long-term trends and patterns within the historical medal data. The model's predictions suggest that the United States, China, and France will continue to dominate the medal standings in the 2028 Olympics, while emerging countries may also achieve significant breakthroughs. The model's ability to predict medal counts for non-medal-winning countries provides valuable insights for national sports committees to plan their sports investment and resource allocation strategies.

The correlation matrix between the number of events and the number of medals (Figure 4) reveals a strong positive correlation, suggesting that countries participating in more events are likely to win more medals. The heatmap visualization (Figure 5) further highlights the efficiency of medal acquisition in different sports, with darker colors representing higher contributions to the total medal count.

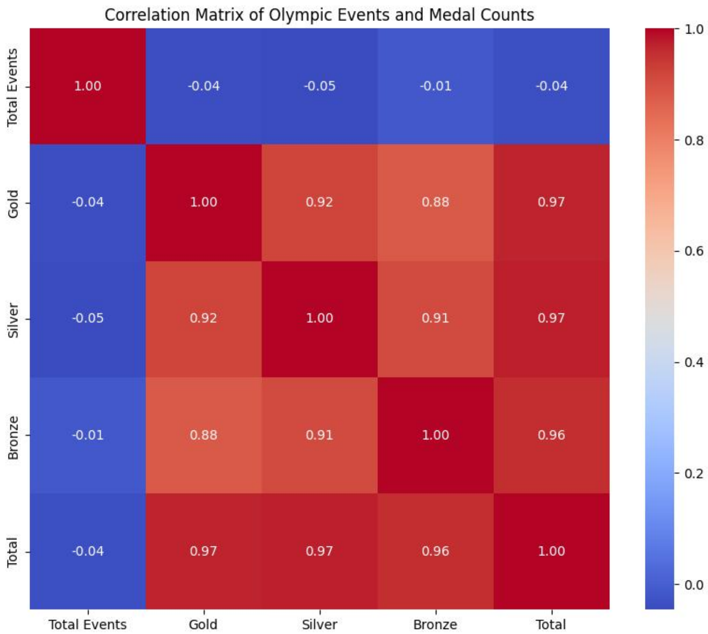


Figure 4 Heat Map of the Correlation Between the Number of Events and Medals

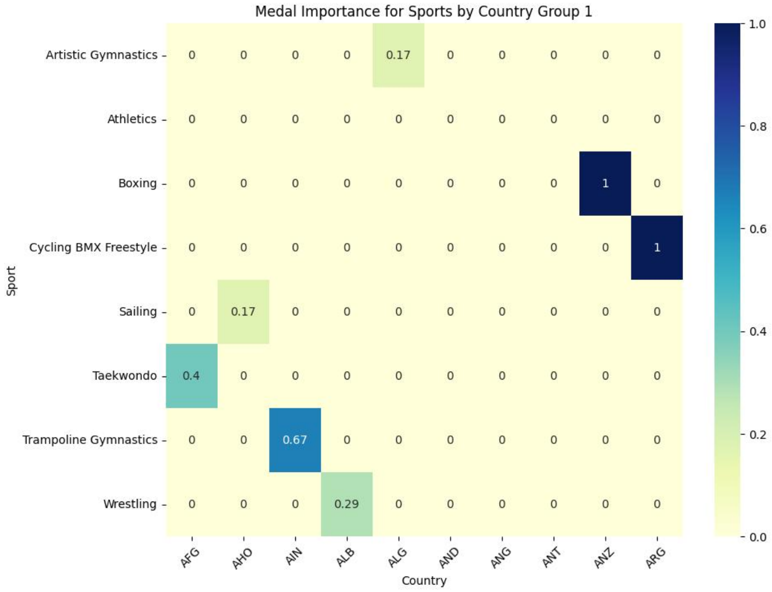


Figure 5 Heatmap of Medal-Acquisition Efficiency in Different Sports

The analysis of the host country advantage (Figures 6, 7, 8) shows that host countries typically win more medals during their hosting year compared to non-hosting years. This phenomenon can be explained by the host country's capacity to expand the number of events and tailor the types of events to suit their strengths.

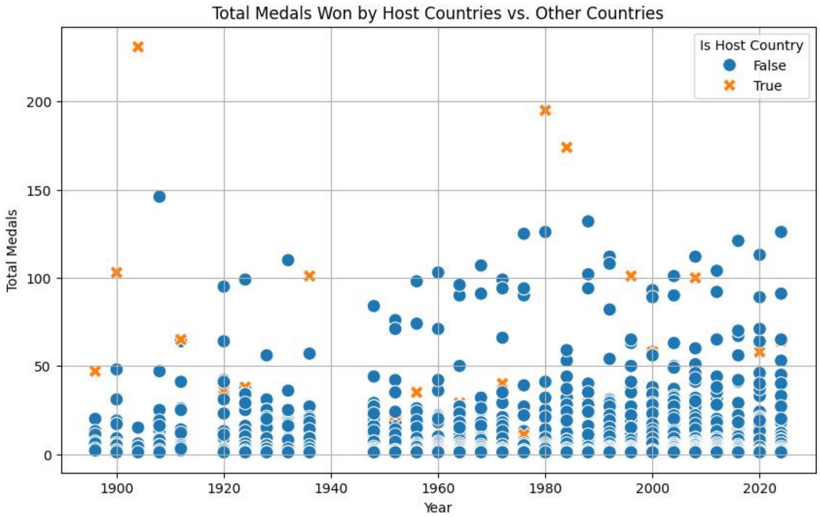


Figure 6 Total Number of Medals Won by Host Countries in Olympic Years

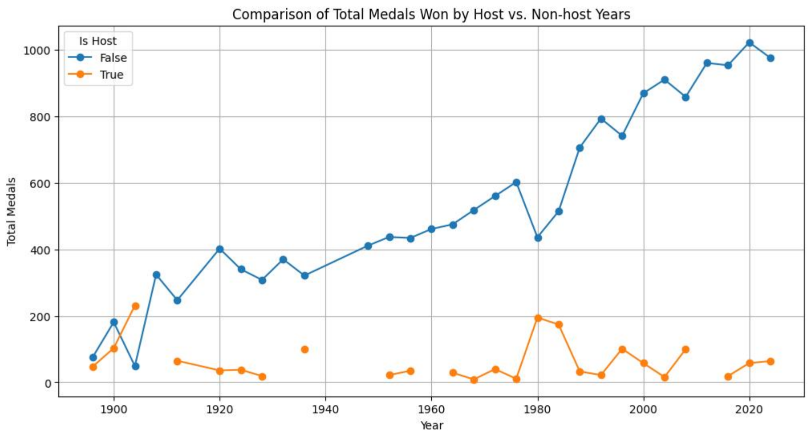


Figure 7 Comparison of Host Country Medal Totals in Olympic vs. Non-Host Years

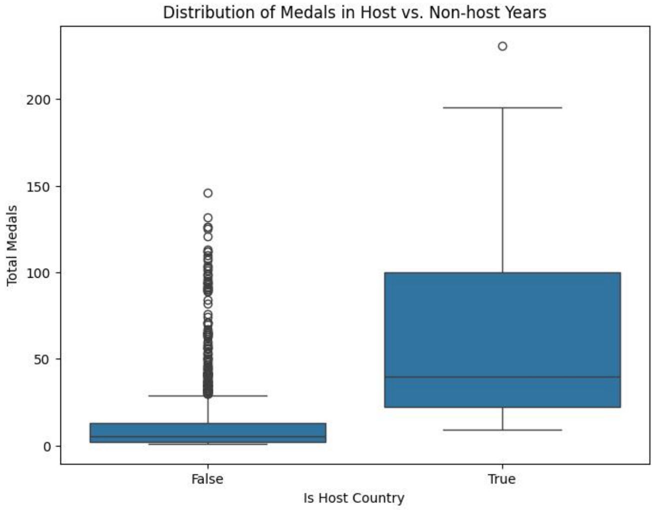


Figure 8 Distribution of Medals of Host Countries in Olympic and Non-Host Years

3.2 Results and Analysis of the Decision Tree Model

The decision tree model was employed to assess the influence of "great coaches" on Olympic medal counts. It was trained using data that included the number of athletes, medal efficiency (medals per athlete), and the medal tally for each country across various sports. The model's structure illustrates how the data is divided based on the number of athletes and medal efficiency to forecast the number of medals.
The model's forecasts (Figures 9, 10, 11) indicate that the introduction of "great coaches" can substantially boost a country's medal count. For example, the model predicts that China could see a substantial increase in medals in sports such as gymnastics and diving if it recruits top-tier coaches.

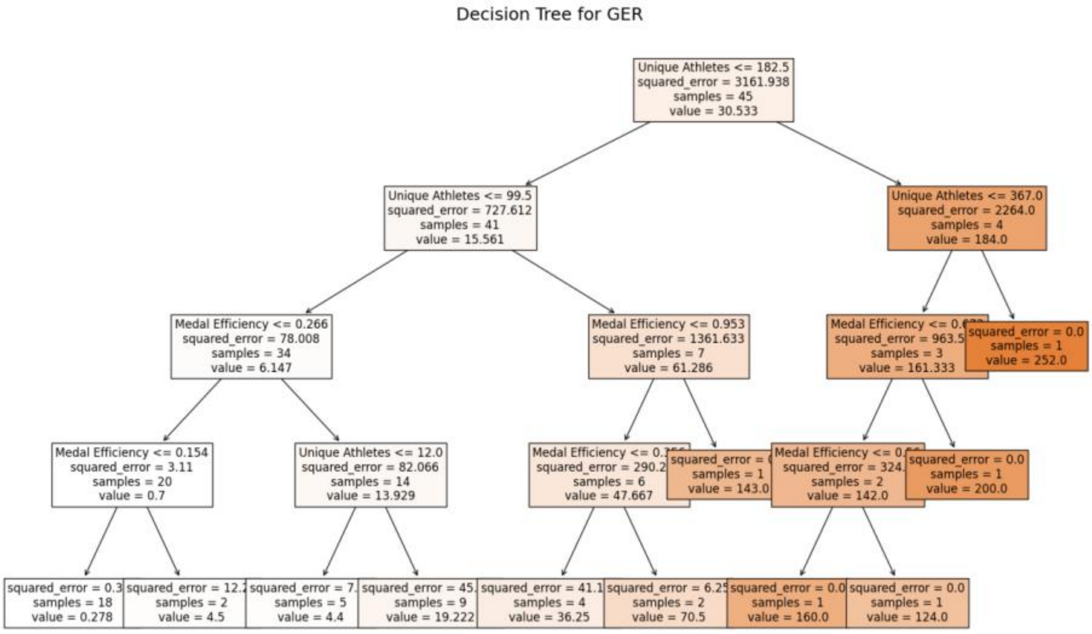


Figure 9 Predicted Medal Counts for Country 1 with and without "Great Coaches"

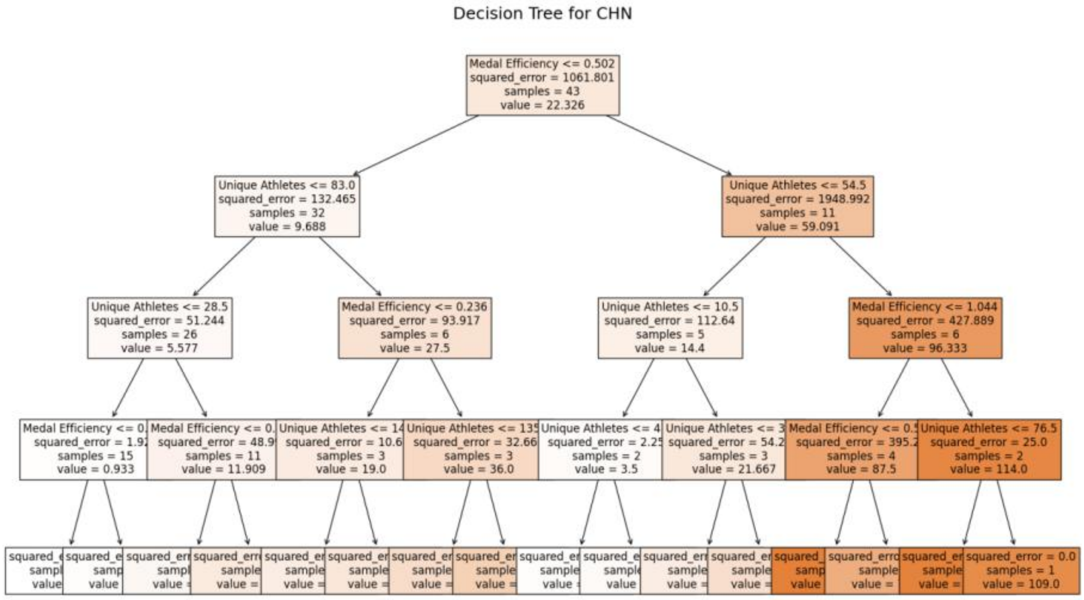


Figure 10 Predicted Medal Counts for Country 2 with and without "Great Coaches"

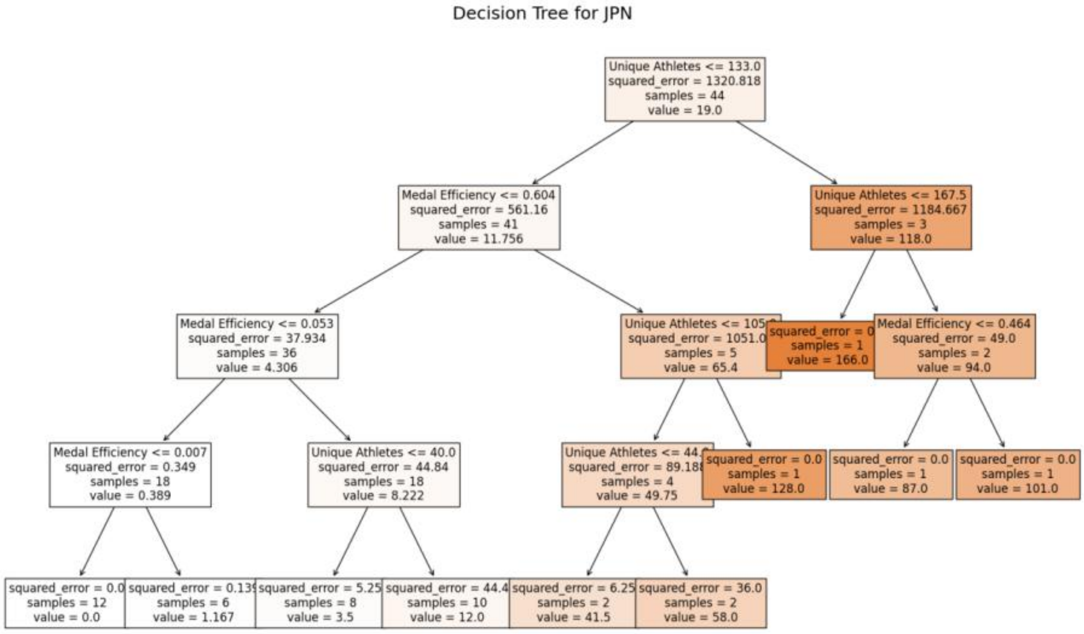


Figure 11 Predicted Medal Counts for Country 3 with and without "Great Coaches"

The decision tree model highlighted that the efficiency of winning medals is a key determinant of a country's overall medal count. The findings suggest that hiring "great coaches" can markedly improve a country's medal performance, especially in sports where the country already possesses a solid base. The model's ability to quantify the impact of "great coaches" provides valuable insights for national sports committees to optimize their coaching resources and strategic planning.

The visualization of the decision tree model shows that the model partitions the data based on the number of athletes and medal efficiency, with the top-level partition being based on medal efficiency. This indicates that medal efficiency is a more important factor than the number of athletes in determining medal counts. The model's forecasts for various countries demonstrate that bringing in "great coaches" can result in a substantial rise in medals, especially in sports where the country already has a high efficiency in winning medals.

To sum up, the decision tree model serves as an effective instrument for examining how "great coaches" influence Olympic medal performance. It also provides useful guidance for optimizing coaching resources and boosting international sports competitiveness.

4 CONCLUSIONS AND OUTLOOKS

The LSTM model was effectively constructed using historical medal data, participating events, and national indicators such as population and GDP, capturing the time-dependence of historical performance and the advantages of sports infrastructure and host country benefits. The model predicted that the United States, China, and France will exhibit strong medal competitiveness at the 2028 Los Angeles Olympics, with potential breakthroughs from emerging nations, providing valuable insights for national sports committees to plan their sports investment and resource allocation strategies. Moreover, the model's ability to predict medal counts for non-medal-winning countries offers a comprehensive understanding of the future distribution of medals, helping to anticipate changes in the ranking of some countries in the medal table. Additionally, the decision tree model quantified the impact of "great coaches" on medal outcomes, revealing that transnational coach mobility significantly influences medal distribution, with notable coaches like Lang Ping and Bela Karolyi enhancing the competitiveness of volleyball and gymnastics, respectively. The model's predictions showed that recruiting top-tier coaches can significantly increase the number of medals won by a country, particularly in sports where the country already has a strong foundation, providing valuable insights for national sports committees to optimize their coaching resources and strategic planning. Furthermore, the study analyzed the correlation between the distribution of Olympic medals and the sports development strategies of each country, indicating that medal distribution is related to sports traditions, advantageous programs, event settings, and factors such as economic level, population size, and sports policies. The analysis of host country advantage showed that host countries typically win more medals during their hosting year compared to non-hosting years, attributed to their ability to increase the number of events and adjust the types of events to their advantage.

Future research may concentrate on improving the predictive accuracy of models by integrating more extensive and current data, such as athlete performance metrics, training methods, and technological innovations in sports. Exploring advanced machine learning techniques, including ensemble methods and deep learning architectures, could further enhance model performance. The results of this study lay the groundwork for optimizing sports resource allocation. Future research could explore specific strategies and policies that national sports committees can implement to maximize their potential for winning medals. The impact of various factors, such as government funding, sports infrastructure development, and athlete welfare programs, can be further analyzed to provide a more holistic approach to sports resource management. The study's insights into the impact of "great coaches" and the host country advantage can inform global sports development initiatives. Future research can explore the role of international collaborations and knowledge sharing in enhancing sports performance across different countries. The study can also serve as a basis for developing policies and programs aimed at promoting sports participation and excellence at the grassroots level, thereby fostering a more inclusive and competitive global sports environment.

COMPETING INTERESTS

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

REFERENCES

[1] Bian X. Predicting Olympic medal counts: The effects of economic development on Olympic performance. The park place economist, 2005, 13(1): 37-44.
[2] Tchamkerten A, Chaudron P, Girard N, et al. Career factors related to winning Olympic medals in swimming. PLoS One, 2024, 19(6): e0304444.
[3] Csurilla G, Fertő I. How to win the first Olympic medal? And the second?. Social Science Quarterly, 2024, 105(5): 1544-1564.
[4] Zheng J, Dickson G, Oh T, et al. Competitive balance and medal distributions at the Summer Olympic Games 1992–2016: Overall and gender-specific analyses. Managing Sport and Leisure, 2019, 24(1-3): 45-58.
[5] Nagpal P, Gupta K, Verma Y, et al. Paris Olympic (2024) Medal Tally Prediction//International Conference on Data Management, Analytics & Innovation. Singapore: Springer Nature Singapore, 2023: 249-267.
[6] Hołub M, Stanula A, Baron J, et al. Predicting breaststroke and butterfly stroke results in swimming based on Olympics history. International journal of environmental research and public health, 2021, 18(12): 6621.
[7] Heazlewood T. Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Journal of Sports Science & Medicine, 2006, 5(4): 480.
[8] De Bosscher V, Shibli S, Weber A C. Is prioritisation of funding in elite sport effective? An analysis of the investment strategies in 16 countries. European Sport Management Quarterly, 2019, 19(2): 221-243.
[9] Deepak V, Anguraj D K, Mantha S S. An efficient recommendation system for athletic performance optimization by enriched grey wolf optimization. Personal and Ubiquitous Computing, 2023, 27(3): 1015-1026.
[10] Thompson F, Rongen F, Cowburn I, et al. The impacts of sports schools on holistic athlete development: a mixed methods systematic review. Sports medicine, 2022, 52(8): 1879-1917.
[11] Wang Y, Wang J, Yang J, et al. STGCN-LSTM for Olympic Medal Prediction: Dynamic Power Modeling and Causal Policy Optimization. arXiv preprint arXiv:2501.17711, 2025.
[12] Jeon M, Lim H. Taekwondo win-loss determining factors using data mining-based decision tree analysis: focusing on game analysis for evidence-based coaching. BMC Sports Science, Medicine and Rehabilitation, 2024, 16(1): 115.