Bayesian Enhanced Stacking Method Integrating Multi-Model Advantages and its Application in Olympic Medal Prediction

Lu Kun[†, a], Yichen Du[†, b], Gaojun Jin[c, *], Wenbo Kuang[d], Jiaxi Li[e]

International College Zhengzhou University, Zhengzhou, China

[a]15236760356@163.com, [b]18860335386@163.com, [c, *]15236760356@163.com, [d]19139935835@163.com, [e]19337960598@163.com

[†]These authors contribute equally.

Abstract

The Olympic medal tally serves as a key indicator for measuring a nation's comprehensive athletic strength, and accurate prediction holds significant importance for sports resource allocation and strategic planning. This study utilizes historical data from 163 countries/regions across Summer Olympics from 1896 to 2024. First, K-means clustering is employed to categorize participating nations into high, medium, and low medal-winning groups based on their performance levels, addressing data heterogeneity. Subsequently, we propose the Bayesian Enhanced Stacking Model (BESM). This model integrates three base learners-LightGBM, Attention-LSTM, and Bayesian Negative Binomial Regression-to capture nonlinear features, long-term temporal dependencies, and count data distribution characteristics, respectively. At the meta-learning layer, Bayesian Linear Regression enables adaptive weight allocation and uncertainty quantification. Systematic comparisons against mainstream models (SARIMA, Prophet, Random Forest, XGBoost, CNN) evaluated performance using MAE, RMSE, NB_Deviance, and SMAPE. Results demonstrate BESM's superiority across all groups: achieving the best performance on at least one metric in 88.5% of cases within the low-tier group (statistically significant, p<0.0001); demonstrated strong overall adaptability in the medium-medal group; and achieved the best SMAPE (33.3%) in the high-medal group. Traditional models exhibited pronounced group dependency and unstable performance. This study validates BESM's effectiveness in handling heterogeneous, sparse sports data, providing an accurate and reliable solution for Olympic medal prediction and national sports strategy formulation.

Keywords

Olympic medal prediction, bayesian ensemble learning, stacking model, time series analysis.

1. Introduction

The Olympic Games, as the world's most influential large-scale sporting event, have consistently drawn global attention to their medal and gold medal standings. The distribution of medals across Olympic Games not only reflects a nation's athletic prowess but also, to a certain extent, mirrors its overall national strength and level of economic development[1]. Therefore, making reasonable and accurate predictions about Olympic medal distributions holds significant value for objectively describing the evolution of the current global sports landscape, assisting nations in formulating strategic plans, and conducting in-depth analyses of the development status of sports programs worldwide[2].

Based on historical medal counts, macroeconomic indicators, sports science factors, and modern data science technologies, current mainstream methods for predicting Olympic medal totals can be broadly categorized into three types: macroeconomic models, historical trend models, and machine learning models. Macroeconomic models focus on regression analysis frameworks that are both data-accessible and highly interpretable (e.g., multiple linear regression [3] and panel data regression [4]). Prediction models based on historical trends exhibit greater diversity, encompassing techniques ranging from moving averages and exponential smoothing to time series analysis [5].

Although the two aforementioned approaches provide valuable baseline predictions for Olympic medals, their predictive accuracy still has room for improvement when addressing scenarios influenced by subjective factors and dynamic changes. To address these shortcomings, researchers have begun shifting their focus toward machine learning models. Reference [6] employed a two-stage random forest [7] to validate the superiority of this model over traditional simple prediction models. Reference [8] conducted a comparative analysis of four models-random forest, BP neural network [9], XGBoost [10], and SVM [11]-on data from the 2028 Los Angeles Summer Olympics, ultimately determining XGBoost as the optimal predictive model.

The predictive accuracy of a single machine learning model is typically constrained by its specific architecture and algorithmic limitations. When addressing complex problems like Olympic medal predictions-characterized by multiple influencing factors, intricate data distributions, and significant heterogeneity-it often proves challenging to balance global stability with local adaptability. To address this, this paper introduces the Bayesian Enhanced Stacking model [12]. This model not only integrates the strengths of different models in capturing temporal trends, handling nonlinear relationships, and adapting to high-dimensional features, but also quantifies prediction uncertainty through its posterior distribution. This enables more robust global inference and more precise local adaptation in the complex and dynamic task of Olympic medal prediction.

The core work of this paper can be broadly summarized in the following three aspects:

1) Preprocess the medal count data from the Summer Olympics for various countries spanning 1896–2024, then employ K-means clustering to categorize participating nations into three tiers based on distinct characteristics.

2) Employed a Bayesian-enhanced Stacking model to forecast medal counts for nations across tiers at the 2028 Los Angeles Summer Olympics. Compared these predictions against five conventional single-model approaches-SARIMAX, Prophet, Random Forest, XGBoost, CNN, and exponential forecasting-to evaluate predictive performance.

3) Based on the research findings, a predictive framework applicable to Olympic medals is proposed, and its practical application value in the field of Olympic medal table forecasting is discussed.

2. Research Methods

Based on data from the Summer Olympic Games medal tables spanning 1896 to 2024, this section outlines the core architecture and fundamental principles of K-means clustering analysis, Bayesian ensemble stacking models, and five comparative models.

2.1. K-means Clustering Analysis

Olympic medal data exhibits significant heterogeneity, reflecting fundamental differences among nations in athletic prowess, resource investment, and historical performance. Historical data reveals that sporting powerhouses like the United States, China, and Russia consistently win dozens or even hundreds of Olympic medals, while numerous developing nations and smaller countries remain stuck at single-digit or zero medal counts. This disparity spanning multiple orders of magnitude makes it challenging for traditional global modeling approaches to simultaneously capture the complex temporal patterns of high-medal-winning nations and the sparse characteristics of low-medal-winning nations.

A single model struggles to maintain optimal performance across all data distributions. By clustering countries with similar characteristics into distinct groups, model parameters and structures can be selected or adjusted for each group to best fit its data properties. This approach enhances the overall performance and robustness of the prediction system at the global level.

Based on the above considerations, this study employs the K-means clustering algorithm to group 163 countries/regions with the following objectives:

(1) To identify and separate groups of nations at different levels of athletic prowess;

(2) To provide a scientific basis for subsequent differentiated modeling;

(3) To enhance the predictive model's adaptability to data heterogeneity;

(4) To improve the interpretability and practical value of the prediction results.

Clustering Results and Grouping Strategy

Through K-means clustering, 163 countries were divided into three tiers:

Tier 1 (High Medal Group): $n_1 = 57$,including traditional sporting powers such as the United States, China, the United Kingdom, Germany, and France, with historical average medals $\mu_{\text{total}} > 20$

Tier 2 (Medium Medal Group): $n_2 = 58$, including emerging sporting nations such as Indonesia, Colombia, and Estonia, with $5 < \mu_{\text{total}} \le 20$

Tier 3 (Low-Winning Group): $n_3 = 48$, including small nations or sports underdeveloped countries like San Marino and Zambia, $\mu_{\text{total}} \le 5$

This grouping serves as the foundation for subsequent modeling. Differentiated feature engineering strategies and model parameter configurations will be applied to each tier to maximize prediction accuracy and model adaptability.

2.2. Bayesian Enhanced Stacking Model

The Bayesian Enhanced Stacking Model (BESM) is an advanced ensemble learning framework based on a two-layer learning architecture. By incorporating Bayesian inference mechanisms, this model theoretically deepens and performance-enhances traditional stacking methods, aiming to address critical challenges in Olympic medal prediction such as strong data heterogeneity, limited sample size, and difficult-to-quantify uncertainty. The core architecture of this model is illustrated in Figure 1:

(1) Overall Framework Design

BESM adopts a hierarchical ensemble strategy, with its core architecture comprising a Base Learner Layer and a Meta-Learner Layer. Let the original training dataset be denoted as $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$ represents the feature vector of the i-th sample, $y_i \in \mathbb{Z}^+$ is the corresponding medal count target value, N is the total number of samples, and d is the feature dimension.

1) First Layer: Base Learner Layer: In the Base Learner Layer, M heterogeneous base models $\{h_1, h_2, \dots, h_M\}$ are constructed, each learning latent patterns in the data from distinct perspectives. This paper selects three types of base learners ($M = 3$):

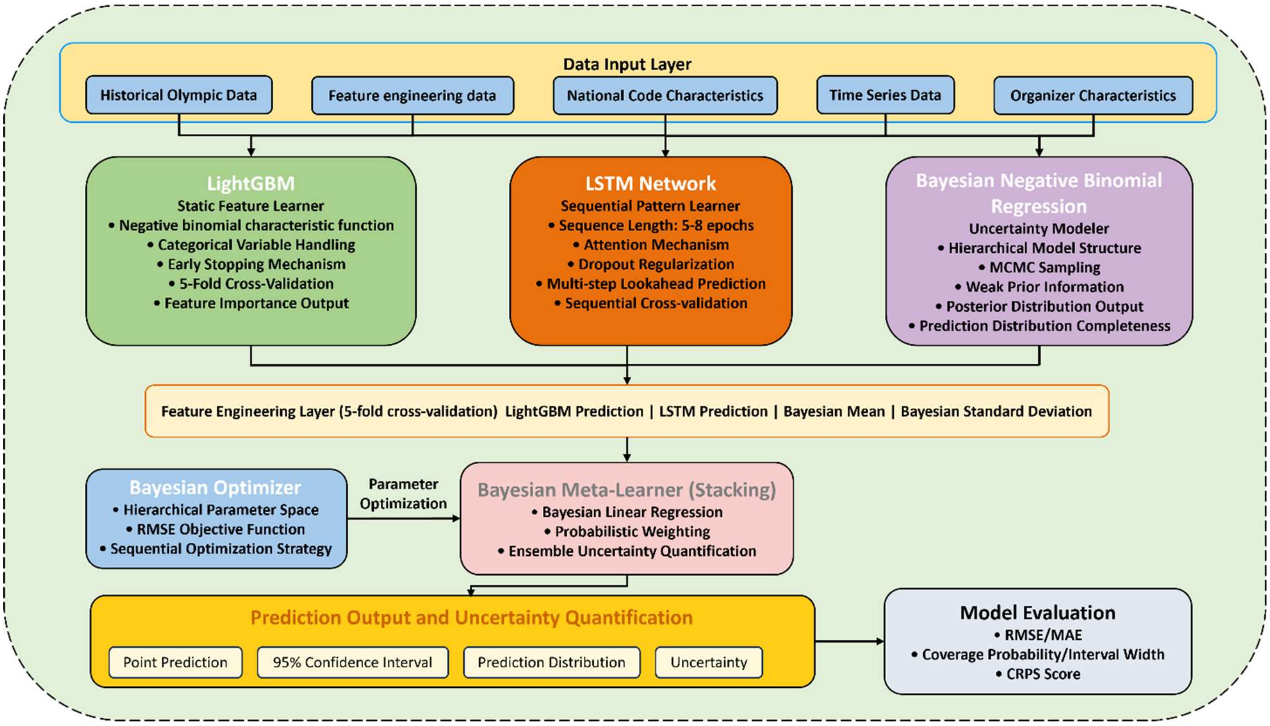

Figure 1. Flowchart of the Bayesian Enhanced Stacking Model

• **LightGBM (h_1):** Captures complex nonlinear interactions among features, excelling in high-dimensional sparse feature spaces;
• **Attention-LSTM (h_2):** Models long-short term dependency patterns in time series by adaptively focusing on critical historical moments via attention mechanisms;
• **Bayesian Negative Binomial Regression (h_3):** Provides a probabilistic prediction framework, quantifying prediction uncertainty and constraining extreme value outputs.

To prevent information leakage, K-fold time-series cross-validation is employed to generate out-of-sample predictions. Specifically, the training set $\mathcal{D}$ is divided into K subsets $\{\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_K\}$ in chronological order. For the kth fold ($k=1,2,...,K$), each base model is trained using $\mathcal{D} \setminus \mathcal{D}_k$ and generates predictions on $\mathcal{D}_k$. After iterating through all folds, the complete out-of-sample prediction matrix is obtained:

$$\mathbf{Z} = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,M} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N,1} & z_{N,2} & \cdots & z_{N,M} \end{bmatrix} \in \mathbb{R}^{N \times M} \tag{1}$$

Here, $z_{i,j} = h_j(x_i)$ denotes the prediction value of the j-th base model for the sample.

2) Second Layer: Meta-Learner Layer: The meta-learner uses the prediction outputs $\mathbf{Z}$ from the base learners as a new feature space to learn an optimal combination strategy. Traditional stacking methods typically employ linear weighting or simple regression models, whereas BESM introduces Bayesian Ridge Regression as the meta-model. Its core advantages include:

①**Adaptive weight allocation:** Automatically learns optimal weights for each base model through posterior inference, eliminating manual parameter tuning;

②**Uncertainty quantification:** Outputs complete prediction distributions rather than point estimates, providing confidence intervals for decision-making;

③**Regularization constraints:** The introduction of prior distributions effectively prevents overfitting and enhances model generalization capabilities.

(2) Mathematical Expression of the Bayesian Meta-Learner

Let the output of the meta-model be a linear combination of the base model predictions:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{M} \beta_j z_{i,j} + \epsilon_i \tag{2}$$

β_0 is the intercept term, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_M]^T$ is the weight coefficient vector, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise.

Bayesian ridge regression constrains the weights by introducing a prior distribution. Suppose the prior for the weights is a zero-mean Gaussian distribution:

$$p(\boldsymbol{\beta}|\alpha) = \mathcal{N}(\boldsymbol{\beta}|0, \alpha^{-1}\mathbf{I}_M) \tag{3}$$

Where α is the precision parameter and $\mathbf{I}_M$ is the M×M identity matrix. The prior for the noise variance follows an inverse Gamma distribution:

$$p(\sigma^2|a,b) = \mathrm{InvGamma}(\sigma^2|a,b) \tag{4}$$

Where a and b are hyperparameters, set in this study as weak prior information (a=1, b=0.1). According to Bayes' theorem, the posterior distribution of the weights is:

$$p(\boldsymbol{\beta}|\mathbf{Z},\mathbf{y},\alpha,\sigma^2) \propto p(\mathbf{y}|\mathbf{Z},\boldsymbol{\beta},\sigma^2) \cdot p(\boldsymbol{\beta}|\alpha) \tag{5}$$

By maximizing the posterior probability (MAP) or solving via variational inference, we obtain posterior estimates of the weights and their covariance matrix, thereby quantifying the uncertainty in the weights. The final predictive distribution is:

$$p(\hat{y}_{\text{new}}|\mathbf{z}_{\text{new}},\mathcal{D}) = \int p(\hat{y}_{\text{new}}|\mathbf{z}_{\text{new}},\boldsymbol{\beta},\sigma^2) \cdot p(\boldsymbol{\beta},\sigma^2|\mathcal{D})\, d\boldsymbol{\beta}\, d\sigma^2 \tag{6}$$

The closed-form solution for this integral is a Student's t-distribution, enabling direct calculation of the predicted mean, variance, and any confidence interval.

(3) Cooperative Mechanisms for Basic Learners

Three types of base learners form a complementary and synergistic relationship within the BESM framework:

1)LightGBM:-Nonlinear Feature Learner: LightGBM accepts complete feature engineering outputs, including lagged features ($y_{t-1}, y_{t-2}, y_{t-3}$), moving statistics (three-period moving average, standard deviation), historical trends (linear regression slope), and host ID ($\mathbb{1}_{\text{host}}$). National codes are treated as categorical variables, and the model loss function adopts a negative binomial distribution to accommodate the over-discrete nature of count data:

$$\mathcal{L}_{\mathrm{NB}}(\theta,r)=-\sum_{i=1}^{N}\begin{bmatrix}\log\Gamma(y_i+r)-\log\Gamma(r)-\log(y_i!)\\+r\log r-(y_i+r)\log(r+\mu_i)\end{bmatrix} \tag{7}$$

where $\mu_i=\exp(f(x_i))$ denotes the predicted mean, r represents the dispersion parameter, and $\Gamma(\cdot)$ is the Gamma function. This loss function is iteratively optimized through gradient boosting, combined with an early stopping strategy to prevent overfitting.

2) Attention-LSTM-Temporal Dependency Modeler: For each country, construct a historical sequence $\{y_{t-T},y_{t-T+1},\ldots,y_{t-1}\}$ of length T as input, where $T\in[5,8]$ is set based on data availability. The hidden state update of the LSTM follows:

$$\begin{aligned}
f_t&=\sigma(W_f\cdot[h_{t-1},y_t]+b_f) \quad (\textit{Forget Gate})\\
i_t&=\sigma(W_i\cdot[h_{t-1},y_t]+b_i) \quad (\textit{InputGate})\\
\tilde{C}_t&=\tanh(W_C\cdot[h_{t-1},y_t]+b_C) \quad (\textit{Candidate Memory})\\
C_t&=f_t\odot C_{t-1}+i_t\odot\tilde{C}_t \quad (\textit{Memory Update})\\
o_t&=\sigma(W_o\cdot[h_{t-1},y_t]+b_o) \quad (\textit{Output Gate})\\
h_t&=o_t\odot\tanh(C_t) \quad (\textit{HiddenState})
\end{aligned} \tag{8}$$

Where $\sigma(\cdot)$ denotes the sigmoid function, $\odot$ represents element-wise multiplication, and W,b is the learnable parameter. The attention mechanism assigns differential weights to historical sequences:

$$\begin{aligned}
e_{t,k}&=v^T\tanh(W_a h_k+U_a s_t) \quad (\textit{Attention Score})\\
\alpha_{t,k}&=\frac{\exp(e_{t,k})}{\sum_{j=1}^{T}\exp(e_{t,j})} \quad (\textit{Normalized Weight})\\
c_t&=\sum_{k=1}^{T}\alpha_{t,k}h_k \quad (\textit{Context Vector})
\end{aligned} \tag{9}$$

The final prediction is mapped through a fully connected layer:

$$\hat{y}_t=W_p c_t+b_p \tag{10}$$

3)Bayesian Negative Binomial Regression-Probabilistic Constrainer: This model employs a hierarchical Bayesian structure to account for cross-country heterogeneity. Let the number of medals won by the gth group of countries follow a negative binomial distribution:

$$y_{i,g}\sim\mathrm{NegBinom}(\mu_{i,g},\phi_g) \tag{11}$$

The log mean modeling is defined as:

$$\log\mu_{i,g}=\alpha_g+\boldsymbol{\beta}_g^T\mathbf{x}_{i,g} \tag{12}$$

The prior settings for group-level parameters are:

$$
\begin{aligned}
&\alpha_g \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2) \\
&\boldsymbol{\beta}_g \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\
&\phi_g \sim \mathrm{Gamma}(a_\phi, b_\phi)
\end{aligned}
\tag{13}
$$

The posterior distribution is obtained through MCMC (Markov Chain Monte Carlo) sampling, with the predicted mean $\mathbb{E}[\mu_{i,g}|\mathcal{D}]$ and standard deviation $\sqrt{\mathrm{Var}[\mu_{i,g}|\mathcal{D}]}$ output as meta-features.

(4) Hyperparameter Optimization Strategy

The hyperparameter space for BESM includes:

1) LightGBM: Tree depth $d \in [3, 8]$, learning rate $\eta \in [0.01, 0.3]$, number of leaf nodes $L \in [20, 100]$;

2) Attention-LSTM: Hidden layer dimension $h \in [32, 128]$, Dropout rate $p \in [0.1, 0.5]$, sequence length $T \in [5, 8]$;

3) Bayesian model: Prior variance $\sigma_\alpha^2, \sigma_\beta^2 \in [0.1, 10]$.

Global search is performed using Gaussian Process-based Bayesian Optimization, with the average RMSE from 5-fold time-series cross-validation as the optimization objective:

$$
\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{K} \sum_{k=1}^{K} \mathrm{RMSE}_k(\boldsymbol{\theta}) \tag{14}
$$

The optimization process employs the Expected Improvement objective function to balance exploration and exploitation, converging to the optimal configuration after 50 iterations.

(5) Mechanism for Quantifying Uncertainty

BESM provides multi-level uncertainty assessment through a Bayesian framework:

a. Epistemic Uncertainty: Arising from insufficient model knowledge, it is quantified by the weighted posterior distribution. For test sample $\mathbf{z}_{\mathrm{new}}$, the variance of the predicted distribution is:

$$
\mathrm{Var}[\hat{y}_{\mathrm{new}}] = \mathbf{z}_{\mathrm{new}}^T \boldsymbol{\Sigma}_\beta \mathbf{z}_{\mathrm{new}} + \sigma^2 \tag{15}
$$

Here, $\boldsymbol{\Sigma}_\beta$ denotes the weight covariance matrix, with the first term reflecting model uncertainty and the second term representing intrinsic noise.

b. Aleatoric Uncertainty: Originating from the inherent randomness of data, characterized by the dispersion parameter ϕ of the Bayesian negative binomial model.

The final prediction output includes:

Through this design, BESM not only delivers precise point predictions but also quantifies forecast reliability, providing comprehensive information support for sports decision-making. This framework demonstrates theoretical advantages and practical value in challenging scenarios such as handling heterogeneous data, small-sample learning, and extreme value prediction.

3. Experimental Results

3.1. Analysis of Experimental Results

(1) Overall Performance Evaluation

This study conducted a systematic model comparison experiment using Olympic medal data from 163 countries and regions spanning 1896 to 2024. Based on historical medal acquisition levels, the sample was categorized into high-achieving groups (Tier 1, n=57), medium-achieving groups (Tier 2, n=58), and low-achieving groups (Tier 3, n=48). Four evaluation metrics were employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Negative Binomial Deviance (NB_Deviance), and Symmetric Mean Absolute Percentage Error (SMAPE). These metrics comprehensively assessed model predictive performance across multiple dimensions, as illustrated in Figure 2.

(2) Robustness Analysis Across Groups

By comprehensively evaluating the results across three groups, BESM demonstrates exceptional robustness across diverse data environments. Among all 18 evaluation metrics, BESM ranked first in 10 dimensions and among the top two in another 6 dimensions. In contrast, other models' strengths were concentrated in specific groups:

1)SARIMA: Demonstrated relatively stable performance in the first-tier group (best MAE/RMSE frequency around 24%), but experienced a sharp decline in the low-reward group (mean MAE 2.6 times that of BESM), reflecting the limitations of traditional time series models when handling sparse count data.

2)Prophet: Competitive in the second-tier groups (best RMSE frequency at 29.3%) due to its trend decomposition capability, but underperforms BESM in low-reward groups (best RMSE frequency only 10.4%).

3)RandomForest and XGBoost: Both slightly outperformed BESM in MAE within the second tier but exhibited insufficient cross-group stability. Although XGBoost achieved the optimal mean NB_Deviance in the third tier, its enormous standard deviation (291.41) revealed sensitivity to outliers.

4) CNN: Deep learning models exhibit exceptional performance on NB_Deviance in the high-ranking group. However, their performance across other groups and metrics is unstable (e.g., best MAE of only 5.2% in the medium-ranking group), and they generally exhibit large standard deviations, indicating a need for more data support.

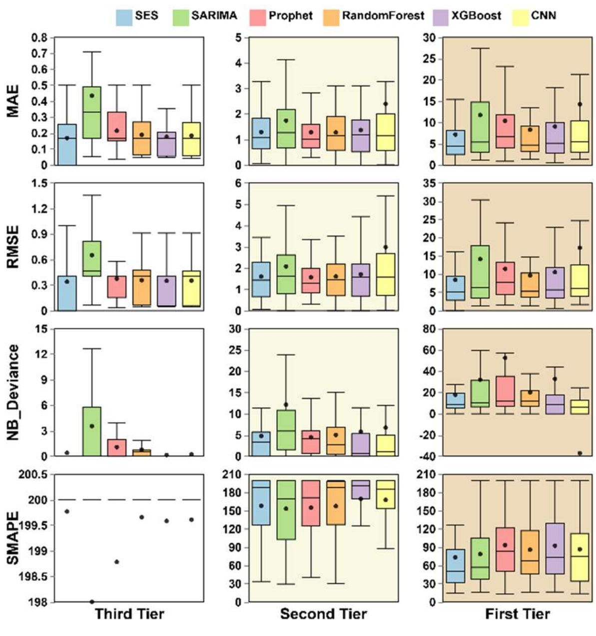

Figure 2. Box Plot of Experimental Results

(3) Advantages of the Bayesian Enhancement Mechanism

The outstanding performance of BESM can be attributed to its multi-level Bayesian enhancement mechanism:

• **Uncertainty Quantification:** Across all groups, BESM exhibits relatively small and stable standard deviations, particularly in low-performing groups (MAE standard deviation 0.18 vs. SARIMA's 0.43), indicating that Bayesian posterior distributions more accurately characterize prediction uncertainty.

• **Adaptive Ensemble Weighting:** By dynamically adjusting base model weights through Bayesian meta-learning, BESM adaptively selects optimal base model combinations tailored to each country's data characteristics. This explains why its best frequency significantly outperforms single models across all groups.

• **Handling sparse data:** In the low-winner group, incorporating Bayesian prior information effectively mitigates overfitting caused by data sparsity, making BESM's advantage most pronounced in this group (average best frequency of 88.5% vs. 7.8% for the next-best model).

3.2. Conclusion

The experimental results fully validate BESM's comprehensive advantages in Olympic medal prediction tasks. The model demonstrates statistically significant and substantial performance superiority in the low-medal-winning countries group ($p<0.0001$, best frequency >85%), maintains robust competitiveness in the high- and medium-medal-winning groups (ranking among the top two in best frequency), and exhibits outstanding cross-data-environment robustness overall. In contrast, other models exhibited pronounced data dependency: traditional time series models (SARIMA, Prophet) suited well-documented high/medium medal-winning nations, while machine learning models (RandomForest, XGBoost, CNN) performed well under specific metrics and groupings but lacked consistency.

4. Summary and Outlook

4.1. Summary

These findings offer significant insights for model selection in sports prediction: when confronted with highly heterogeneous and sparsely populated real-world datasets, the Bayesian ensemble learning framework delivers more reliable and robust predictive solutions by integrating the strengths of multiple models and quantifying prediction uncertainty.

4.2. Future Outlook

Although this study has achieved certain results, several directions warrant further exploration in future work:

Enrichment and Optimization of Feature Systems: This study primarily relies on historical medal data. Future research could incorporate broader feature variables, such as annual GDP, population structure, sports funding, athlete development systems for specific events, and even sociocultural factors from various countries. This would establish a multi-source, heterogeneous feature system to more comprehensively reveal the driving mechanisms behind medal distribution.

Extension of Application Scenarios: The core concepts and methodology of this framework can also be transferred to other predictive scenarios with similar data characteristics (such as temporal, heterogeneous, and sparse data). Examples include forecasting outcomes for other major international events (e.g., Winter Olympics, World Cup) or applying it to socio-economic domains for demand forecasting and risk assessment.

References

[1] Bernard A B, Busse M R. Who wins the Olympic Games: Economic resources and medal totals[J]. Review of economics and statistics, 2004, 86(1): 413-417.

[2] **ang B, Zhao B. Olympic medal analysis and prediction[C]//Proceedings of the 2025 5th International Conference on Applied Mathematics, Modelling and Intelligent Computing. 2025: 765-772.

[3] Guo Y F. FORECASTING OLYMPIC MEDAL COUNTS: A MULTIPLE LINEAR REGRESSION MODEL[J]. 2025.

[4] Zulfikar R, STp M M. Estimation model and selection method of panel data regression: An overview of common effect, fixed effect, and random effect model[J]. JEMA: Jurnal Ilmiah Bidang Akuntansi, 2018, 9(2): 1-10.

[5] Wang Z, Yang W. Olympic Medal Prediction Based on Multiple Regression and Time Series Analysis[J]. *IJHSA*, 2025, 46(4): 6491-6502.

[6] Schlembach C, Schmidt S L, Schreyer D, et al. Forecasting the Olympic medal distribution during a pandemic: a socio-economic machine learning model[J]. arxiv preprint arxiv:2012.04378, 2020.

[7] Wu X, He J, Yip T, et al. A two-stage random forest method for short-term load forecasting[C]//2016 IEEE Power and Energy Society General Meeting (PESGM). IEEE, 2016: 1-5.

[8] 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.

[9] Li J, Cheng J, Shi J, et al. Brief introduction of back propagation (BP) neural network algorithm and its improvement[M]//Advances in computer science and information engineering: volume 2. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 553-558.

[10] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.

[11] Wang H, Hu D. Comparison of SVM and LS-SVM for regression[C]//2005 International conference on neural networks and brain. IEEE, 2005, 1: 279-283.

[12] Zhang C, Chen M. Prediction of Undrained Shear Strength Utilizing a Hybrid Stacking Model Enhanced by Bayesian Optimization Algorithm[J]. Transportation Research Record, 2025, 2679(3): 457-471.