Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 11
TPCEE 2025

Multi-Objective System Predicting Based on Hybrid LSTM-Markov and GSRF Models

Yichen Lin *

University of Electronic Science and Technology of China, Chengdu, China

* Corresponding Author Email: lyc490415@gmail.com

Abstract. The prediction of multi-objective complex systems has always been one of the significant challenges in academia. The multi-dimensional structure, diverse distribution, and complex temporal dependencies of the data make it difficult for conventional mathematical models to perform effective analysis. This study explores the application of hybrid machine learning algorithms in predicting multi-objective systems. This paper developed an LSTM-Markov model which uses the Markov chain to optimize the preliminary predictive results obtained by LSTM. A Grid Search-optimized Random Forest (GSRF) model was integrated with the LSTM-Markov model to address the zero-inflated and long-tailed characteristics of the chosen datasets. The GSRF model was a powerful non-linear classifier that could resolve the zero-inflated outputs. Using Olympic medal counts prediction as the case, the final evaluation metrics of LSTM-Markov revealed that the MSE and R-squared value of the LSTM-Markov model were 0.12 and 0.89 respectively, indicating significant improvements upon baseline methods. The classification accuracy of GSRF achieved 95.52%, demonstrating promising performance in classification and forecasting. This study provides an innovative approach to solving the multi-target predicting problems of complex time series systems. The method also holds significant application potential in areas such as financial risk prediction and disease transmission modeling.

Keywords: Multi-objective system; LSTM, Markov chain; Zero-inflation; Grid Search Random Forest.

1. Introduction

A Multi-Objective Complex System refers to a dynamic system composed of multiple interrelated and interdependent objectives or subsystems, exhibiting non-linearity, high dimensionality, uncertainty, and emergent behavior [1]. Such systems are prevalent in natural, engineering, and social domains, requiring simultaneous optimization or coordination of multiple objectives. The modeling and predicting of these systems can be extremely challenging. Therefore, research in related fields has become increasingly important.

Many different approaches are being used by various researchers and scholars for the purpose of projecting and modeling multi-objective systems. Nagpal et al. [2] have used multiple linear regression to predict the medal counts. They selected various socioeconomic indicators to determine the linear expressions of the dependent outputs, achieving an $R^2$of 0.98. Schlembach et al. [3] have applied a two-staged Random Forest algorithm to a multi-dimensional medal dataset. They first employed a Random Forest as a classifier to determine whether a nation should win any medal at all, and then estimate the exact number of medals. In the paper by Zhao [4], they have established a GA-BP algorithm model, combined with Genetic Algorithm (GA) and backpropagation neural network (BPNN), to optimize the weights and bias parameters of the BP neural network. Utilizing the global search capability of genetic algorithm, the training efficiency and prediction performance were greatly improved. In the paper by Wang [5] which is titled “An ensemble learning based prediction strategy for dynamic multi-objective optimization”, they proposed an ensemble learning based prediction strategy (ELPS) to help algorithms re-initialize a new population after a change is detected. They use the four base prediction models in ELPS, including knee point-based autoregression model (KP-AR), population-based autoregression model (P-AR), linear prediction model (LP) and random re-initialization model (RND). The four base models will be trained by the historical information with ensemble learning once a change happens, and then a strong prediction model can be constructed on these four models. The results revealed that the ELPS has better performance in handling dynamic multi-objective optimization problems, compared with other cutting-edge prediction strategies on benchmark test suite. Wang et al. [6] have proposed a novel hybrid model, STGCN-LSTM, to forecast medal distributions by integrating the spatio-temporal relationships among countries and the long-term dependencies of national performance. The Spatial-Temporal Graph Convolution Network (STGCN) captures geographic and interactive factors such as coaching exchange and socioeconomic links while the Long Short-Term Memory (LSTM) module models historical trends in medal counts, economic data, and demographics. A Zero-Inflated Compound Poisson (ZICP) framework is incorporated to separate random zeros from structural zeros, providing a clearer view of potential breakthrough performances.

However, forecasting the behavior of variables in multi-objective systems remains challenging despite recent methodological and computational advances. Traditional single-objective linear regression and Grey Prediction models struggle to capture the nonlinear characteristics of systems and the diversity of data structures. While machine learning and deep learning algorithms, which have gained prominence in recent years, perform well in the analysis and prediction of some complex systems, they often face challenges such as limited potential for improvement, poor interpretability, and low generalizability. Other issues include the lack of high-quality open-source datasets and a unified platform to compare different data preprocessing methods and prediction models [7]. The paper proposes a novel algorithm based on an LSTM-Markov and GSRF hybrid model, which maintains high accuracy in prediction tasks for multi-objective complex systems with imbalanced data structures. Compared to traditional baseline models and machine learning approaches, the proposed method reduces the regression prediction error on complex time series systems to MSE = 0.09. Additionally, when using GSRF to address the zero-inflation problem, the accuracy reaches 95.52%. In conclusion of the key contributions, this paper introduces a highly feasible and accurate hybrid machine learning model, particularly suitable for multi-objective prediction scenarios in complex time series systems, and it provides a reference for future innovations in related fields.

2. Methodology

To analyze the patterns of multi-objective complex systems, we first need to collect a huge volume and wide diversity of data. Thus, we created four datasets from the medal counts raw database to undertake data mining and feature engineering. The first dataset contains the Olympic medal counts from 1896 to 2024. The second dataset includes the participating athletes together with their nationalities, gender, specific events, and final awards from 1896 to 2024. The third dataset records the information about the official arrangements of relevant events from 1896 to 2024. The final dataset enumerates the host cities and their respective host countries for each game from 1896 to 2024. To resolve the zero-inflated outputs caused by the severe imbalance in athletic strength between nations, the Grid Search Random Forest Model was established to deal with the disparity between countries that consistently earn medals and those never have one. For countries that often appears on the medal table, like the U.S, China, and some European countries, the LSTM-Markov model was developed, since the corresponding data can be treated as time-series. Identifying the potential breakthrough countries was addressed by apply the GSRF model as a classifier, since the given problem can be seen as whether a leap from 0 to 1 will be achieved.

2.1. LSTM-Markov Model

To predict the medal counts of multiple countries as a multi-objective system, we had considered several methods. By observing the dataset, we discover complex temporal dependencies and nonlinear characteristics of the data structure. Therefore, several advanced deep learning models were considered, such as Transformer, LSTM and GRU [8]. Unlike traditional neural networks such as RNN, which struggles with vanishing gradient problems and can only learn short-term relationships, LSTM combines long-term and short-term memory to ensure the stability and accuracy of the model along the time series [8,9]. While Transformers have shown promise in long-sequence tasks, their quadratic computational complexity and need for large datasets are impractical for our medium-size medal database. LSTM provides a balanced trade-off between performance and computational efficiency. However, it often introduces lag issues, so we incorporate a Markov model to correct the predicted results of each sequence. We evaluate the model performance by measuring metrics including Mean Squared Error (MSE) and R2 score. The architecture of our model is illustrated in Figure 1.

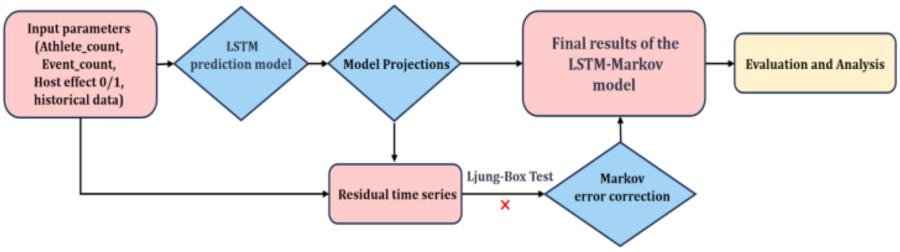

Figure 1. workflow of LSTM-Markov Model

LSTM - Long Short-Term Memory Networks, proposed in 1997 [8,10], are a unique class of RNN which has better processing capability for gradient disappearance in long sequence data training. The working principle of LSTM is shown in Figure 2.

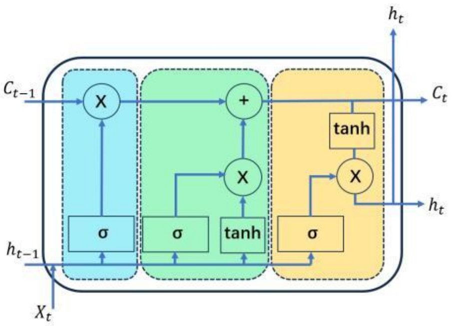

Figure 2. The diagram of LSTM

LSTM can recognize long-term dependencies in the time-series data by using memory gates. These gates determine the state of the machine learning algorithm. LSTM has two transitive states, $c_t$and h_t, which improve the processing efficiency compared to RNN's one state. An LSTM cell contains a forget gate $f_t$an input gate i_t, and an output gate o_t. The calculation process of LSTM can be expressed as:

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

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

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{4}$$

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

$$h_t = o_t \odot \tanh(c_t). \tag{6}$$

Where $\sigma(\cdot)$ represents the Sigmoid function, W_f,W_i, W_c, and W_o represent weighted matrix, b_f, b_f,b_f, and b_f represent the corresponding bias, and $\odot$is the Hadamard Product, which refers to the elementary multiplication of the corresponding entries in the matrices [8,9,10].

After bringing the input parameters into our LSTM model, we obtain the predicted values of medal counts in 2028. To verify if the precision of the results is satisfactory, we conduct a Ljung-Box white noise test [11], whose formula is:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \tag{7}$$

Where n is the sample size, m is the number of lags, $\widehat{\rho_k}$ is the sample autocorrelation coefficient at lag k. The p value of the Ljung-box test was examined and found to be 0.03, leading to the rejection of the null hypothesis. The residuals are not completely white noise, indicating that LSTM has not fully captured the long-term dependencies or some underlying state transition patterns in the data. To further improve the precision of the prediction, we introduced the Markov chain for error correction [12].

For ease and clarity, we present the workflow of the Markov correction on the total medal count of the United States in 2028. We construct a state transition matrix based on the prediction errors of the U.S. medal counts from 1948 to 2024. The frequencies of the prediction error states for the first 19 Olympic Games are used as the initial state vector. By multiplying the initial state vector with the state transition matrix, we obtain the probability of the 2028 U.S. total medal prediction error being in each state. The state with the highest probability is selected as the error state for the predicted total medal count of the United States in 2028, which is then used to adjust the forecasted value. Thoroughly, the state transition matrix is constructed with the following steps.

Firstly, the relative error e_i for each predicted value was calculated, using the obtained values and real values of total medal counts of the U.S. in each Olympic Games from 1948 to 2024 (denoted as $M_{p,i}$ and $M_{t,i}$, $i = 1,2,3,\ldots 19$):

$$e_i = \frac{M_{p,i} - M_{t,i}}{M_{t,i}} \tag{8}$$

Secondly, we applied the K-means clustering algorithm to divide the 19 relative errors into three distinct ranges, as shown in Table I. These are referred to as three different state spaces, $[\mathrm{Q1}, \mathrm{Q2}, \mathrm{Q3}]$ respectively.

Table 1. Markov State Interval Division

Symbol	State	Interval
Q1	Normal State	[0, 0.15]
Q2	Overestimated State	[0.15, 0.3]
Q3	Extremely Overestimated State	[0.3, 0.5]

Thirdly, each relative error e_i is assigned to one of the three states $[\mathrm{Q1}, \mathrm{Q2}, \mathrm{Q3}]$ based on the cluster it falls into. The frequency of transitions between these states is used to build the transition matrix P as shown in Equation 9, in which the elements represent the probability of transitioning from one state to another between two Olympic Games.

$$P = \begin{pmatrix} 0.2857 & 0.5714 & 0.1429 \\ 0.3333 & 0.5556 & 0.1111 \\ 0 & 1 & 0 \end{pmatrix} \tag{9}$$

The entries of the state transition matrix P, denoted as $P_{i,j}$, represent the probability of one step transitioning from state Q_i to state Q_j which can be calculated by:

$$P_{i,j} = \frac{N_{i,j}}{N_i} \tag{10}$$

Where N_i is the number of occurrences of state Q_i, and $N_{i,j}$ is the frequency of state Q_i transitioning to state Q_j in one step.

Once the state transition matrix was established, we build the initial state vector X_0 using the frequency P_i, which is the occurrence of state Q_i:

$$X_0 = [0.316, 0.579, 0.105] \tag{11}$$

Then we combined the transition state matrix P to predict the error state vector X_1:

$$X_1 = X_0 \cdot P \tag{12}$$

The state with the highest probability in X_1 was selected as the error state for the 2028 U.S. medal prediction. The clustering center corresponding to this state was employed as the relative error value e_c for the sample point to be corrected. Thus, the corrected final result of the projection M_c was obtained:

$$M_c = \frac{M_p}{1 + e_c} \tag{13}$$

Where M_p is the initial output by LSTM which have not been processed. The above operation was repeated and applied to other countries that have won at least one Olympic medal. It is worth noting that the total medal count M_p and gold medal count G_p predicted by LSTM were both corrected and then used to form the medal table.

2.2. Grid-Search Random Forest Model

When predicting Olympic medal counts, a common challenge is the disparity between countries that have won medals and those that have never won any [13]. Countries without previous medals are recorded as zero in data processing, resulting in an excessive number of zeros in the dataset. This phenomenon is academically referred to as the zero-inflation problem. To address zero-inflated outputs, several methods have been proposed by researchers and practitioners, such as the Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB). However, these models lack the capability to handle large-scale complex datasets in the context of Olympic medal count prediction. Worse still, they perform poorly in nonlinear problems. Therefore, to address this problem, it was transformed into a binary classification task, which would be solved by the Grid Search Random Forest (GSRF) model [14]. We analyzed the number of countries that broke their medal drought in each game and used this data to train the GSRF model, which can predict how many countries will achieve a breakthrough in the next game.

GSRF - The Grid Search Random Forest is an advanced variant of the Random Forest model with its hyper-parameters optimized by Grid Search algorithm. Grid Search operation utilizes exhaustive search over all possible combinations in a given parameter search space. Following this, cross-validation is conducted to evaluate each combination, and then the configuration yielding the highest validation score (e.g., AUC, F1-score, or MSE) will be chosen as the terminal hyperparameters. Random Forest is a classic machine learning bagging algorithm that uses decision trees as estimators [15]. It integrates multiple decision trees by randomly selecting data subsets with replacements while randomly choosing a subset of features as input. To generate multiple training subsets, it uses bootstrap sampling to generate multiple training subsets, with each tree trained on a different subset. A baseline Random Forest model was trained without hyperparameter tuning to evaluate its initial predictive performance. In classification tasks, Random Forest takes the majority voting result as the final output. In our model, it incorporates cross-validation to prevent over-fitting, reducing the cost and errors. The block diagram of our GSRF method is depicted in Figure 3.

Feature Engineering - The first step of GSRF is choosing the indicators to be selected as input features for the model. Various factors influencing the achievement of medals were considered, including the number of consecutive participations, the growth rate of participants, and the number of events participated in during the past five games. Finally, three indicators were determined to be fed into the model: consecutive participation factor, events participation rate and new events factor.

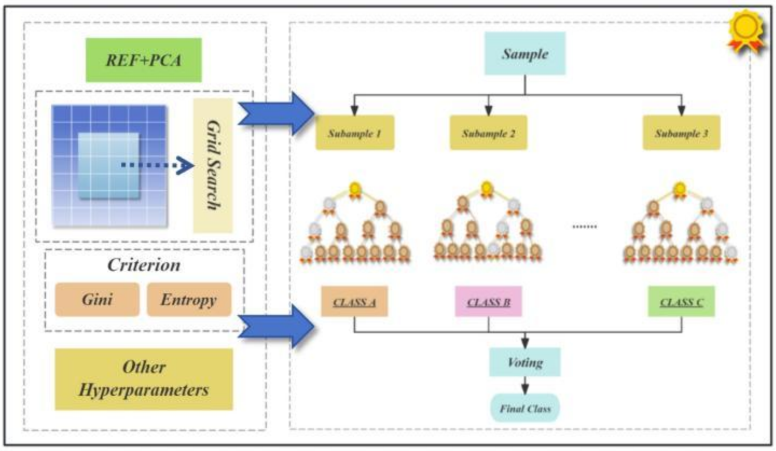

Figure 3. Block diagram of GSRF

To implement the GSRF model, the dataset was split into a training set and a test set in an 8:2 ratio. Utilizing the GridSearchCV algorithm in the sklearn.model_selection module [16], the hyperparameter space was specified and fully searched, RandomForestClassifier algorithm from the scikit-learn (sklearn) machine learning library's ensemble module was then used to initialize the Random Forest model. By fitting the Grid Search targets on the training set, all the feasible hyperparameter combinations were automatically iterated, and the optimal combination was then determined, which is presented in Table II. It is worth noting that the 5-fold cross-validation was performed to assess the outcome of each combination, which can also prevent over-fitting on the training set. Moreover, it balanced evaluation reliability and training data volume, as a 5-fold CV allocates 20% samples per fold, suitable for small and medium-size datasets.

Table 2. Optimal Hyperparameters for GSRF

Parameter	Values
max_depth	20
min_samples_leaf	4
min_samples_split	10
n_estimators	400

3. Results and Discussion

3.1. Results of LSTM-Markov model

To demonstrate the strong predictive and analytic capabilities of the proposed hybrid ML model for multi-objective complex systems, the Olympic medal counts was selected as the target system due to its complex temporal nature, multiple dependencies, and imbalanced dataset. Based on the implementation of the LSTM-Markov method, results predicting the total medal counts and gold medal counts of 2028 Olympic Games are yielded. After 500 training iterations of LSTM, a preliminary predictive result of medal table was derived. The model is fine-tuned by applying hyper-parameter tuning, varying number of epochs, neurons, dropout rate and the layers of the underlying neurons. Then, by using the Markov chain, prediction intervals with higher accuracy were acquired. Due to the space limitations, this study selects the United States and France as representatives and presents the trends in the number of gold medals and total medals using line charts. A medal table featuring the top seven countries in total medal counts has been generated. It contains number of gold medals and total medals for each country in 2028, along with the corresponding confidence interval at the 95% confidence level and changes compared to 2024. Evaluation metrics used for the model include MSE, RMSE, MAPE and R2 Score [17].

Below figures 4 and 5 show the forecast model using LSTM and Markov for the trend of gold medals and total medals for the United States. As it can be seen from both figures the U.S. is likely to see a significant increase in gold medals and total medals in 2028. Another fact is that there was a pronounced peak in 1984, which is closely linked to the effect of the host country on the United States that year.

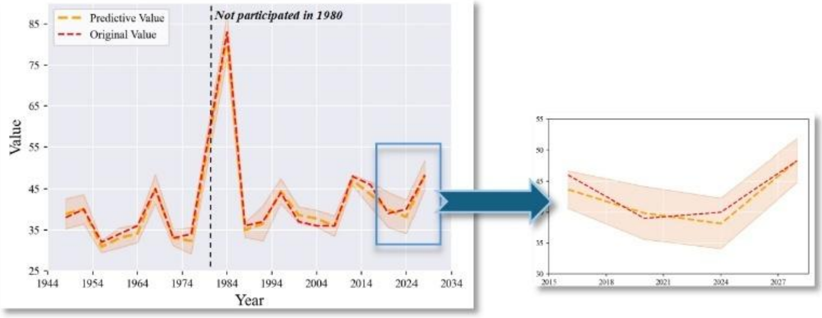

Figure 4. Forecasting gold medals of U.S. using LSTM-Markov

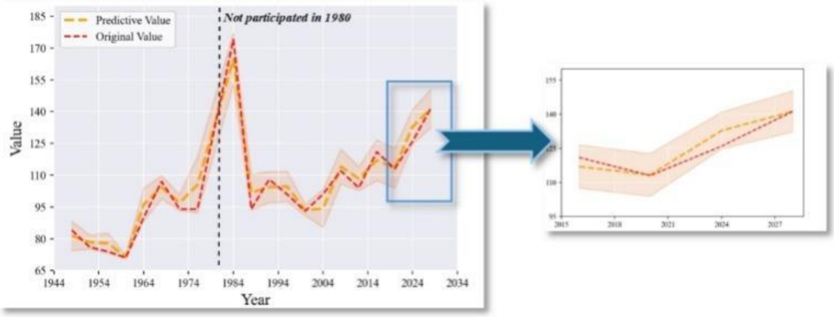

Figure 5. Forecasting total medals of U.S. using LSTM-Markov

Below figures 6 and 7 show the forecast model for the trend of total medals and gold medals for France. A sharp decline from 2024 to 2028 can be seen from both figures, which is probably associated with the turnover of the host effect. The oscillations and peaks are closely related to the evolution of Olympic events and whether the high-level French athletes were scheduled to compete. If some outstanding athletes (such as those who have won Olympic medals or strong emerging talents) were scheduled to compete, their probability of winning an Olympic gold medal would be significantly higher than that of average athletes. It is important to note that in this study, socioeconomic factors were not used as input parameters. Instead, predictions were made based on historical data and relevant event parameters, such as the number of participating athletes and the number of events involved.

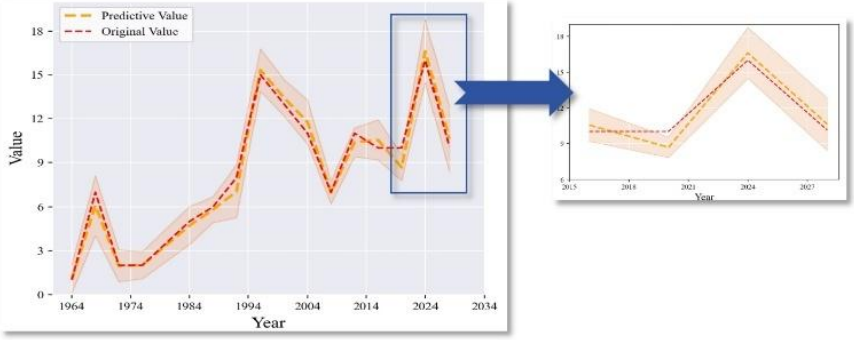

Figure 6. Forecasting gold medals of France using LSTM-Markov

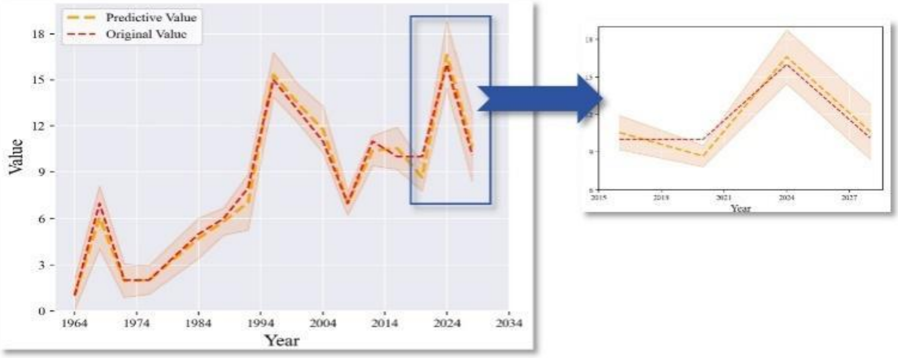

Figure 7. Forecasting total medals of France using LSTM-Markov

The predictive medal table for 2028 Los Angeles Olympic Games was produced by repeating the same process as mentioned (United States and France), which is displayed in Table III. Besides U.S., countries like Japan and Korea will perform better in 2028, while China, Germany and Australia seem to undergo a fall. These changes can be attributed to the mobility of athletes and the influence of varying events. The number of medals in some countries' dominant sports may be reduced or increased, while some emerging sports are often popular only in a few countries, thereby altering the medal counts to some extent. For instance, skateboarding was officially included as an Olympic event in 2020 Tokyo Olympics, and Japan achieved great success in relevant events, securing multiple medals and leveraging its home advantage.

Table 3. Predictive Medal Table for 2028 Los Angeles Olympics

Medals / Country	Gold medals in 2028	Gold medals interval	Total medals in 2028	Total medals interval	Trend
USA	49	[46,52]	137	[132,142]	↗
CHN	36	[34,38]	84	[80,88]	↘
JPN	21	[19,23]	47	[44,50]	↗
AUS	16	[15,17]	50	[46,54]	↘
FRA	10	[9,11]	43	[39,47]	↘
KOR	14	[12,16]	34	[31,37]	↗
GER	10	[8,11]	26	[24,28]	↘

To evaluate the precision of the final projections, several commonly used metrics are employed, including MSE, RMSE, MAPE and R2 Score. The formula of these metrics are as follows.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{17}$$

Where n is the number of samples, y_i and $\hat{y}_i$ are the real and predictive values of total medals or gold medals respectively, $\bar{y}$ is the arithmetic mean of y_i. By substituting the acquired values into the respective formulas, the performance metrics were determined as follows: $MSE = 0.09$, $RMSE = 0.30$, $R^2\ score = 0.92$,$MAPE = 28.3\%$. Compared with the results by simply applying traditional LSTM ($MSE = 0.12$, $R^2\ score = 0.89$) ， the LSTM-Markov model demonstrates superior prediction accuracy and performance. Additionally, the LSTM-Markov model exhibits better interpretability and enhanced feature extraction capability.

3.2. Results of GSRF model

Based on the training of the GSRF classifiers, results enabling the identification of potential new Olympic medal-winning countries are yielded. Evaluation metrics used for the model included Precision, Recall, Accuracy, F1-score and ROC-AUC [18]. Below figure 8 depicts the confusion matrix heatmap for the 2024 Paris Olympic prediction, where the columns represent predicted class 0 (no breakthrough) or 1 (winning the first medal), and each row represents the real class (0 or 1) the country belongs to.

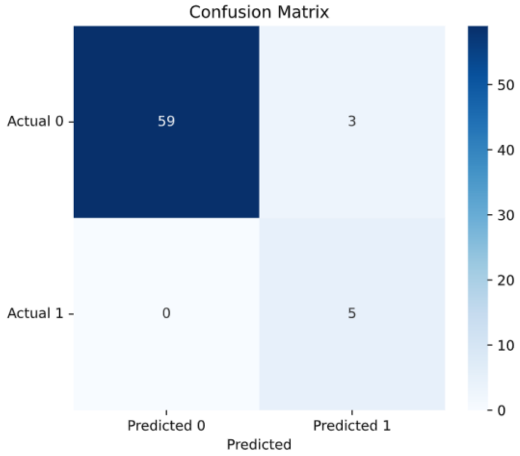

Figure 8. Confusion matrix of GSRF

From the matrix, it is clear that the model correctly identified 59 negative case (True Negative, TN) and 5 positive cases (True Positive, TP). However, three False Positives were observed, indicating that the model predicted three countries as potential first-time medal winners, while none of the actual negative cases were missed (i.e., FN=0). Thus, it is predicted that there would be 5 countries breaking their medal drought (TP + FN = 5 + 0 = 5).

To further analyze the model's predictive ability, several evaluation metrics as mentioned above were determined:

$$Precision = \frac{TP}{TP + FP} = \frac{5}{5+3} = 62.50\% \tag{18}$$

$$Recall = \frac{TP}{TP + FN} = \frac{5}{5+0} = 100\% \tag{19}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 76.92\% \tag{20}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 95.52\% \tag{21}$$

Since the Recall equals 100%, it suggests that the model is highly effective at identifying first-time medal-winning nations, making it particularly useful for scenarios where missing such predictions would be costly. However, the model yielded a moderate Precision of 62.50%, which implies that 37.50% of predicted new winners were False Positives. This could stem from overfitting to noisy training data or insufficient features (e.g., lacking socioeconomic or recent athletic performance indicators). These drawbacks could be addressed through further model refinement, additional feature engineering, or incorporation with more diverse training data. The F1-score (76.92%) balances the two metrics, reflecting a robust trade-off between Precision and Recall. Additionally, the overall accuracy (95.52%) demonstrates high correctness in predictions across all categories (both new and non-new winners). The ROC curve is drawn as figure 9. It can be observed that the curve tends to approach the point (0,1), and the area under the ROC curve (AUC) is 0.92, indicating that the GSRF model effectively separates the two classes. Finally, this model was applied to the prediction of the 2028 Los Angeles Olympics, yielding the result that four countries will win their first Olympic medal.

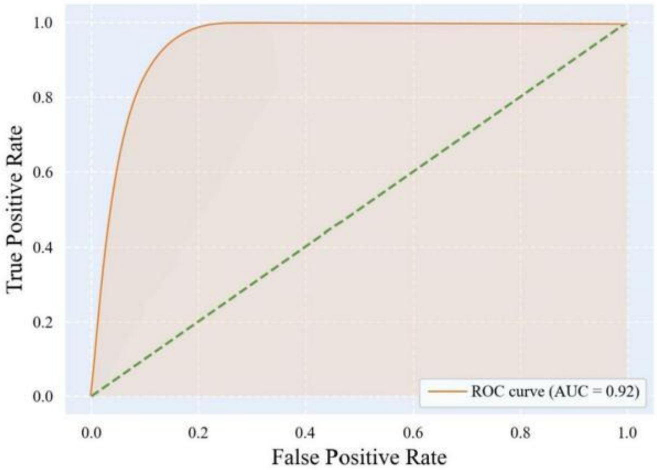

Figure 9. ROC curve of GSRF

Comparison of the proposed method with Benchmark methods is presented in Table IV with respect to the metrics evaluated. Overall, the results obtain in the study demonstrate that the GSRF model is a promising tool for forecasting Olympic medal breakthroughs, with strong predictive accuracy and reliable generalization capabilities. In the upcoming work, advanced feature engineering techniques tailored to Olympic data, such as incorporating dynamic features (e.g., annual athlete performance trends, funding changes) and using feature selection methods like Recursive Feature Elimination (RFE), could remove redundant or noisy features as well as reduce FP by capturing more discriminative patterns. These techniques provide valuable insights into the zero-inflated problem in Olympic Games prediction, enabling researchers to design more targeted model and algorithms to produce more reliable forecasts.

Table 4. Comparison of The Proposed Method with Benchmark Methods

Method	Accuracy	Precision	Recall	F1-score
Gradient Boosting Machines	0.912	0.720	0.785	0.751
Support Vector Machine	0.901	0.741	0.768	0.754
Naive Bayes	0.885	0.647	0.742	0.691
Decision Trees	0.893	0.730	0.752	0.741
K-Nearest Neighbors	0.872	0.725	0.720	0.722
GSRF(proposed method)	0.955	0.625	1.000	0.769

4. Conclusion

The study demonstrates the effectiveness and superiority of using hybrid Machine Learning algorithms to perform prediction and modeling of multi-objective complex systems, especially for those with complicated time series dependencies and diverse data structure such as zero-inflation and long-tailed distribution. By leveraging the LSTM-Markov model, temporal features of the medal database can be extracted and dynamic residual corrections can be performed, providing feasible insights of handling the complicated dependencies between multiple variables for relevant sports professionals and researchers. The implementation of the GSRF classification model addresses the zero inflation problem in the database, namely the disparity between countries that have won medals and those that have not, which further enhance the hybrid model's capability of dealing with uncertainty and non-linearity. Through a case study, the results highlight the feasibility of integrating advanced Machine Learning and Deep Learning algorithms to significantly improve prediction accuracy on multi-objective complex systems. This research also offers valuable insights that such methods could be extended to other areas such as economics, finance, and environmental fields, showing strong generalizability and broad application potential.

Looking ahead, optimizing the GSRF classification algorithm through threshold adjustment and data imbalance handling holds promise for enhancing Precision and minimizing False Positive Rate. Further hyperparameter tuning of the LSTM-Markov model may improve its effectiveness in complex time-series prediction. Incorporating dynamic features and exploring ensemble learning techniques could contribute to greater prediction stability and accuracy. To enhance the interpretability of the above Machine Learning algorithms, in the future work, attention mechanisms can be incorporated to clearly observe which time steps or features the model focuses on during prediction. Additionally, feature importance analysis methods, such as Permutation Importance, SHapley Additive exPlanations (SHAP), and Local Interpretable Model-Agnostic Explanations (LIME), can be utilized to precisely quantify the impact of input features on the results. This approach breaks through the black-box nature of LSTM models, allowing for a clearer understanding of how the model makes decisions using different input features.

References

[1] Zheng Y, Wang D X. A survey of recommender systems with multi-objective optimization[J]. Neurocomputing, 2022, 474: 141-153.

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

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

[4] Zhao S, Cao J, Steve J. Research on Olympic medal prediction based on GA-BP and logistic regression model[J]. F1000Research, 2025, 14: 245.

[5] Wang F, Li Y, Liao F, et al. An ensemble learning based prediction strategy for dynamic multi-objective optimization[J]. Applied Soft Computing, 2020, 96: 106592.

[6] Wang Y, Wang J, Yang J, et al. STGCN-LSTM for Olympic Medal Prediction: Dynamic Power Modeling and Causal Policy Optimization[J]. arXiv preprint arXiv:2501.17711, 2025.

[7] 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.

[8] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, 2019, 31(7): 1235-1270.

[9] DiPietro R, Hager G D. Deep learning: RNNs and LSTM[M]//Handbook of medical image computing and computer assisted intervention. Academic Press, 2020: 503-519.

[10] Landi F, Baraldi L, Cornia M, et al. Working memory connections for LSTM[J]. Neural Networks, 2021, 144: 334-341.

[11] Hassani H, Yeganegi M R. Sum of squared ACF and the Ljung–Box statistics[J]. Physica A: Statistical Mechanics and its Applications, 2019, 520: 81-86.

[12] Stewart W J. Introduction to the numerical solution of Markov chains[J]. 2021.

[13] Wang Y, Wang J, Huang T Y, et al. STGCN-LSTM for Olympic Medal Prediction: Dynamic Power Modeling and Causal Policy Optimization[J]. arXiv preprint arXiv:2501.17711, 2025.

[14] Ma X, Wang Z. Performance and Score Analysis of Athletes Based on EWM-TOPSIS Algorithm and GSRF Prediction Model[C]//2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE). IEEE, 2024: 933-937.

[15] Genuer R, Poggi J M, Genuer R, et al. Random forests[M]. Springer International Publishing, 2020.

[16] Ahmad G N, Fatima H, Ullah S, et al. Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV[J]. Ieee Access, 2022, 10: 80151-80173.

[17] Prakash S, Jalal A S, Pathak P. Forecasting covid-19 pandemic using prophet, lstm, hybrid gru-lstm, cnn-lstm, bi-lstm and stacked-lstm for india[C]//2023 6th International Conference on information systems and computer networks (ISCON). IEEE, 2023: 1-6.

[18] Kumar D, Kothiyal A, Kumar R, et al. Random Forest approach optimized by the Grid Search process for predicting the dropout students[C]//2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET). IEEE, 2024: 1-6.