International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1357-1367

https://doi.org/10.70517/ijhsa463105

Sports big data-driven prediction of elite athletes' competitive status fluctuation combined with time series analysis

Xu Yan[1] and Wei Hao[2,*]
[1] Physical Education Department, North China Electric Power University, Baoding, Hebei, 071003, China
[2] Safety Work Office, Xingtai Medical College, Xingtai, Hebei, 054000, China
Corresponding authors: (e-mail: xtyzbwc@163.com).

Abstract Sports big data-driven combined with time series analytics greatly improves the prediction of competitive status fluctuations in elite athletes. In this study, an athletic state prediction framework incorporating Informer-based time series analysis is constructed based on multi-source sports big data. The model is utilized to learn the physiological data characteristics of athletes in different competitive states. The prediction performance of the model is specifically analyzed by combining the root mean square error and other evaluation indexes. Athletic state case prediction of 30 elite athletes is realized by deep learning feature data. The physiological data collected in the study show that athletes in the best competitive state generally have body temperatures between 36.5°C and 38°C, and other indicators are maintained in a relatively normal range. The model in this paper achieves good prediction results in different athletic states, with F1 values above 0.94 and prediction errors between 0 and 0.2. The model has a small error value in performing physiological data prediction, such as the absolute error of body temperature is between 0~0.5℃, which realizes the accurate prediction of elite athletes' competitive state.

Index Terms Big data-driven, Time series analysis, Informer model, Athletic state prediction

I. Introduction

With the continuous development of technology, data collection and analysis of athletes has become a new hot spot [1], [2]. The wide application of big data technologies has changed many fields, and these technologies can also be used in the field of sports health to provide more accurate predictions of fluctuations in athletic status and make recommendations for athletes, especially elite athletes who are the mainstay of the sport [3]-[5].

Big data techniques can convert vectors and matrices into numerical values and then convert these values into valid eigenvectors or eigenmatrices [6], [7]. These feature vectors or feature matrices are analyzed to extract feature information to derive predictions of the athlete's state [8]. The forms of data collection of athletes include sensor measurements, sports competitions, psychological questionnaires and health profiles, i.e., traditional medical physical examination data, athletes' own technical data, etc., can be used for data analysis [9]-[11]. And big data in the prediction of fluctuating competitive status of elite athletes include injury prediction, performance prediction and emotional state prediction [12], [13]. By collecting and analyzing the information of athletes' training load, psychological condition, health records and sensor data, the possible injury and disease conditions of athletes can be predicted in advance, so that corresponding countermeasures can be taken to maximize the protection of athletes' physical health [14]-[17]. Through the information of athletes' technical data, sports career records, psychological conditions, and sensor data, it is possible to develop a reasonable training program and predict the performance status of athletes [18], [19]. This helps to improve the training effectiveness of athletes and helps coaches to develop the best tactics to win the game [20], [21]. And the influence of emotional state on athletes' performance is even more negligible [22]. By analyzing the data of athletes' psychological condition, mood fluctuation, and sleep quality, we can predict the emotional state of athletes and help athletes to maintain the best state of mind to improve the performance of training and competition [23]-[25].

Literature [26] constructed a system based on big data (BD) technology to analyze the state of high-level neural activity of athletes, and based on the BD spectral clustering algorithm by evaluating the learning efficiency of the system, data collection, etc., the effectiveness of the system was determined, and it can improve the efficiency of athlete training. Literature [27] proposed a sports injury prediction model based on BD and Artificial Intelligence (AI), combining deep learning and Random Forest algorithm to analyze the physiological and psychological factors of athletes in order to improve the accuracy of injury prediction, and verified the effectiveness of the model, which significantly reduces the injury rate. Literature [28] examined the evaluation of sports effectiveness based on feature selection, designed a sports evaluation system using the client-server model, and proposed the general architecture of the system. The effectiveness of the method effectively improved the efficiency and accuracy of sport effect evaluation. Literature [29] explored the application of Support Vector Machines (SVMs) in the prediction of professional sports injuries and used Big Data Analytics (BDA) techniques to provide useful insights, proving the effectiveness of the SVM model with very high accuracy and prediction rates. Literature [30] elaborated that big data effectively reduces the rate of injuries in athletes with a wide range of applications and it does so by predicting the health status of athletes for use in identifying possible future sports injuries. Literature [31] used fuzzy theory calculation method based on information security to install relevant testing equipment on athletes, showing that the equipment is able to effectively predict and adjust athletes' competitive status through data collection in order to adjust training programs and training time.

In this paper, we borrowed the Informer time series analysis model to construct a framework for predicting the competitive status of elite athletes. The Encoder-Decoder architecture therein is adopted to capture the long dependencies in the sequences. The adaptability of the self-attention mechanism is improved by unifying the input representation, and the time complexity problem is optimized using ProbSparse Self-attention to further improve the prediction performance. The framework is used to analyze the characteristic data such as body temperature, blood pressure, heart rate, blood oxygen, etc. of athletes in different competitive states, and mine the correlation between physiological data and athletes' competitive states. Finally, the prediction effect of the proposed model is evaluated by combining the actual prediction cases of elite athletes' athletic status.

II. Athletic performance prediction from time-series analysis of fused multi-source data in sport

II. A. Informer-based time series forecasting models

Time series forecasting [32] is a regression method based on historical data to predict future changes, the principle of the method is to consider the development of things in continuity, through the statistical analysis of data in the past time, to infer the future development trend.

Informer [33] improves the shortcomings of the traditional Transformer [34] by reducing the quadratic computational complexity of self-attention in time series prediction, the bottleneck of memory usage when stacking multi-layer networks, and the slow speed of step-by-step decoding prediction.

II. A. 1) Encoder-Decoder Architecture

The deep information processing flow embedded in the encoder-decoder paradigm used by Informer is shown in Figure 1. Using a step-by-step process of dynamic decoding, the input information representation X^t is encoded into the hidden state H^t and the corresponding output representation is generated by the decoder. In this process, the decoder computes the state h_k^t from the previous step and the other necessary outputs from the kth step, generates a new hidden state h_{k+1}^t, and predicts the sequence of the (k+1)th step. The advantage of this architecture is its flexibility to adapt to input sequences of different lengths and complexities. The dynamic decoding mechanism allows the decoder to continuously adjust its state during the processing of the sequence to better capture long dependencies in the sequence.

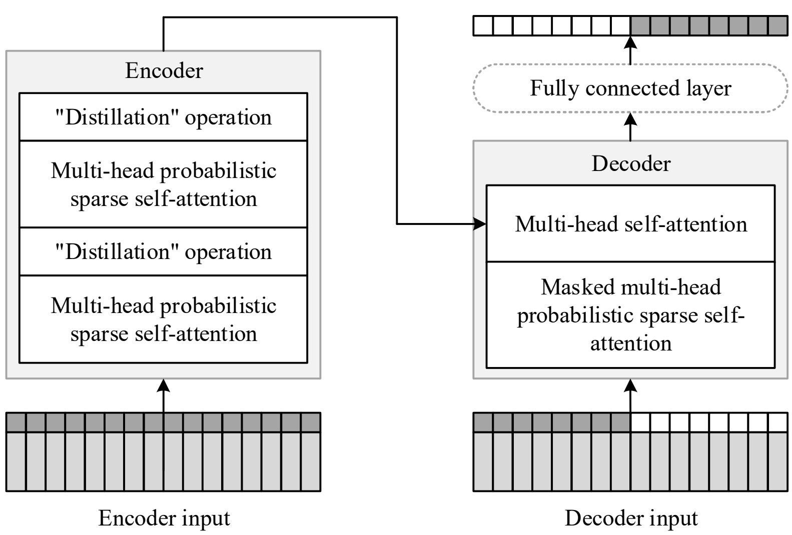

Figure 1: Basic Architecture Diagram of the Informer model

II. A. 2) Harmonization of input representation

RNN [35] based models do not rely on timestamps and capture time series patterns through the loop structure itself, and the original Transformer uses the dot product self-attention mechanism and employs timestamps as encoding of location information. However, in long time series prediction problems, temporal modeling requires not only local temporal information, but also hierarchical timestamps (week, month, year) and agnostic timestamps (holidays,events). Conventional self-awareness mechanisms are difficult to adapt directly, and query-key mismatch between encoder and decoder degrades the prediction performance. Therefore, Informer gives a unified input representation as shown in Figure 2.

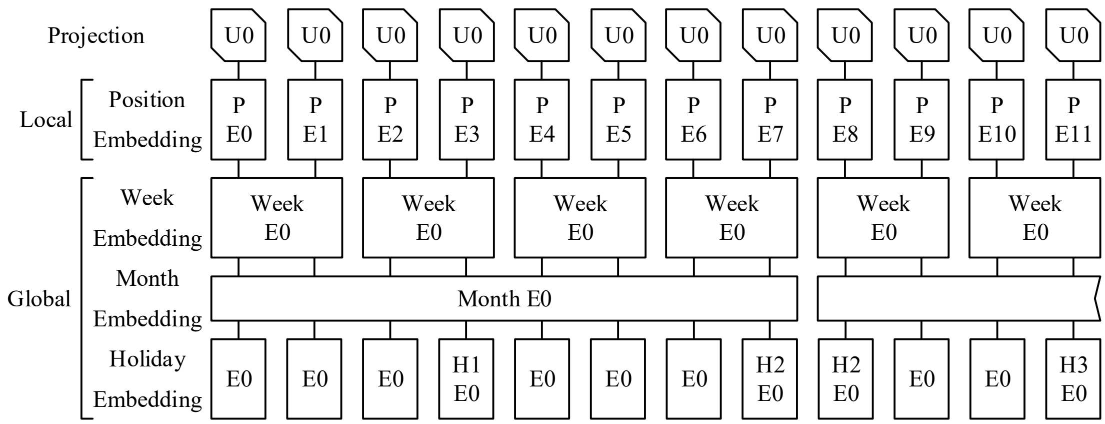

Figure 2: Unified Input representation Diagram

II. A. 3) ProbSparse Self-attention

ProbSparse Self-attention is used to solve the time complexity problem of Self-attention dot product computation by optimizing the time complexity from $O(L^2)$ to $O(L\log L)$, by sampling $\log L$ points for dot product computation instead of selecting the whole L to be dot product.

Attention feature maps are of two types: activation-based attention maps and gradient-based attention maps, the difference between the two being whether or not an activation function is used.

In the attention feature maps, notation $A^m \in R^{C_n \times H_n \times W_n}$, A_m represents the mth layer of the network's activation output, where C_m, H_m and W_m denote the number of channels, height and width, respectively. Generating a graph that visualizes the internal working mechanism of the network is essentially a search for a mathematical mapping relationship, as in Eq:

$$R^{C_\infty \times H_\infty \times W_\infty} \to R^{H_\infty \times W_\infty} \tag{1}$$

The absolute value of each element in the mapping directly reflects the significance of its contribution to the final output of the entire network. Accordingly, we construct the described mapping function by calculating the absolute values of these elements across channel dimensions and further analyzing them in a statistically significant way.

There are many redundant combinations of V-values in the feature maps of the encoder in Informer, therefore, in order to select superior features with dominant characteristics, this paper employs a distillation operation and generates a self-attentive feature map with focus in the next layer.

The distillation is computed in Eq:

$$X'_{j+1} = MaxPool(ELU(Con1d([X_j^t]_{AB}))) \tag{2}$$

Here $[\cdot]_{AB}$ denotes the attention block containing the multi-headed ProbSparse self-attention and other basic operations. To improve the robustness of the distillation operations, Informer employs a strategy of halving the inputs to build a copy of the main stack, and gradually reduces the number of self-attention distillation layers, discarding one layer at a time to keep their output dimensions aligned. As a result, this study joins the outputs of all the stacks together to form the final hidden representation of the encoder.

II. A. 4) Generative Decoder Decoder

The Decoder used by Informer is different from traditional Decoder, generative Decoder generates all the predicted outputs at once, unlike selecting a specific flag as a token, Informer selects a sequence as long as L_{token} from the input sequence as a token.

The decoder generates the long sequence output, using generative inference to mitigate the speedup. Using vectors as input to the decoder see Eq:

$$X_{de}^{t} = concat(X_{token}^{t}, X_{0}^{t}) \in R^{(L_{odon}+L_{y})\times d_{modol}} \tag{3}$$

Here $X_{tokem}^{t} \in R^{L_{salem}\times d_{modcl}}$ is the start marker, and $X_{0}^{t} \in R^{L_{y}\times d_{modcl}}$ placeholder is used to represent the target sequence. Masked multi-head attention is applied, in which by setting the mask point to -∞, each position is able to capture its subsequent position information prospectively, thus effectively circumventing the limitations imposed by the autoregressive problem. These features incorporating forward and backward dependencies are transformed by the fully connected layer to form the output of the model, whose output size d_y depends on whether univariate or multivariate prediction is performed.

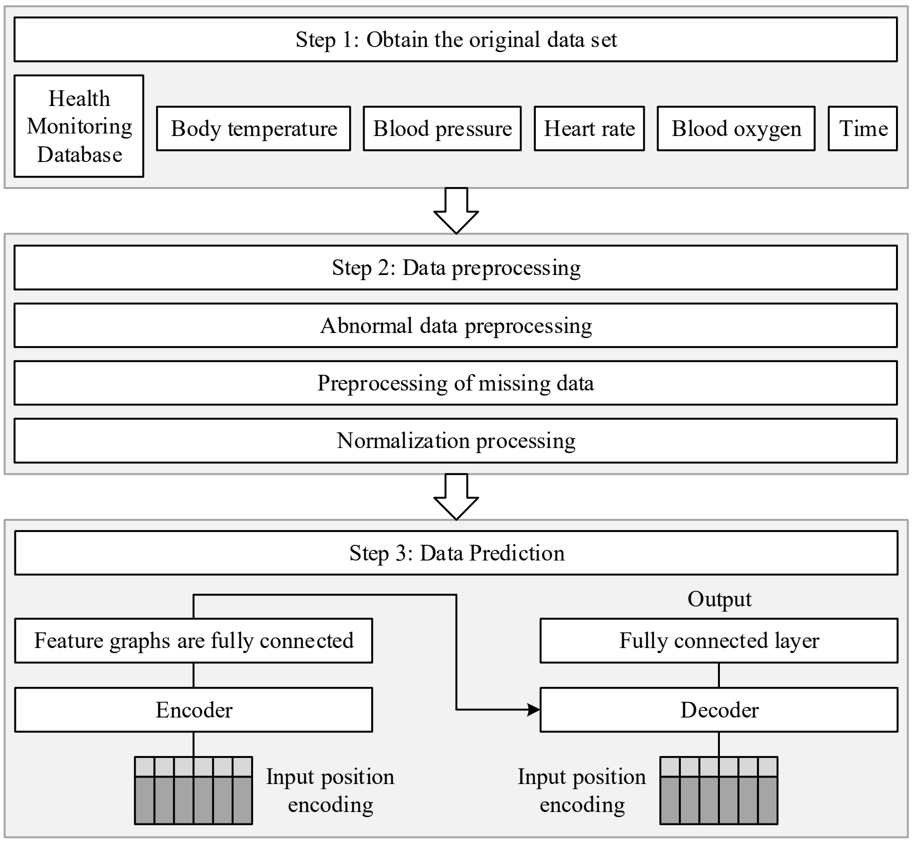

Figure 3: Overall Framework of the prediction model

II. B.Athletic state fluctuation prediction model based on time series analysis

In recent years, research progress has been made on physiological data prediction models for athletes' competitive status. Most of these models are mainly constructed based on techniques such as CNN and LSTM, and trained using existing publicly available datasets to predict single physiological indicators such as heart rate and blood pressure of individuals. However, there are obvious limitations of this approach. First, since these models usually predict only for a single indicator, they cannot take into account the interactions and effects between different physiological data. Second, these models perform poorly in terms of practicality, accuracy, and generalization ability, making it difficult to meet the complex and changing practical needs.

In this section, by elaborating the multidimensional data input and specific applications and feature extraction methods, we aim to solve the limitations and problems in the current research on physiological data prediction of athletes' competitive status, and propose a multidimensional physiological data prediction model based on the Informer model, which is able to consider the correlation between each physiological data and simultaneously predict the body temperature, blood pressure, heart rate, and oxygen, which are the four dimensions of physiological data. By learning the physiological data characteristics of athletes in different competitive states, the accurate prediction of athletes' competitive states is realized.

II. B. 1) Overall framework of the forecasting model

The overall framework of the Informer prediction model is shown in Fig. 3 with the following steps:

(1) Acquisition of the original dataset. In this study, feature data including body temperature, blood pressure, heart rate, blood oxygen, etc. were extracted and organized from the database.

(2) Data preprocessing. Firstly, the abnormal data in the dataset were identified and transformed into vacant data; secondly, the vacant data were preprocessed; finally, all the data were normalized.

(3) Data prediction. The processed data are divided into training set, testing set and validation set, which are used as inputs to the prediction model, and are trained in several iterations to get the prediction results. In this study, 80% of the experimental data is used as a training set for model training, 10% is divided into a test set to evaluate the performance of the model, and the remaining 10% is used as a validation set for model optimization and validation.

II. B. 2) Data pre-processing

Data preprocessing refers to the manipulation, cleaning, integration, and transformation of raw data in order to improve data quality, usability, and applicability. In this study, the following preprocessing was performed on the experimental data:

(1) Abnormal data processing. Abnormal data refer to those data that do not conform to the regular data distribution or do not conform to the expected data patterns. To identify these anomalous data, the mean and standard deviation of the data set are first calculated, and then a threshold range is determined based on the mean and standard deviation. If a data point is outside of this threshold range, it is considered anomalous data. For the processing of abnormal data, the abnormal data are first deleted directly, and then the values are added using the missing data processing method.

(2) Missing data processing. In order to ensure that the prediction study is carried out properly, the study used the Lagrange interpolation method to process the missing data. The method is based on the idea of Lagrange function, which solves for the values of unknown data points by interpolating polynomials from known data points. The formula for the Lagrange interpolation method is as follows:

$$L_n(x)=\sum_{i=0}^{n-1} y_i p_i(x) \tag{4}$$

$$p_i(x)=\prod_{\substack{i=0\\ i\neq j}}^{n} \frac{x-x_i}{x_j-x_i} \tag{5}$$

where y_i is the value of the function at a known point and $p_i(x)$ is the Lagrange fundamental polynomial corresponding to the i th value point.

(3) Normalization. In order to unify the scale, prevent small data from being swallowed, and improve the training efficiency of the prediction model and enhance the accuracy of the model prediction results, the study adopts the Z-score normalization method for the dataset that has been subjected to the anomalous and vacant data processing, which transforms the data into a distribution centered on the mean and measured by the standard deviation, and realizes the normalization of the data. The formula for the normalization process is shown in Eq:

$$z=\frac{x-\mu}{\sigma} \tag{6}$$

where x represents the original data and μ represents the average of all the data, indicating the average value obtained by adding up all the data and then dividing by the total number. σ represents the standardized variance, which is used to measure the degree of dispersion of the data, and a larger standardized variance indicates a greater volatility of the data. z denotes the standardized value, which converts the raw data into a degree of deviation relative to the mean by subtracting the mean and dividing by the standard deviation.

II. B. 3) Error evaluation indicators

In order to verify the prediction effect of the model, the experiment uses three error evaluation metrics to assess the accuracy of the model's prediction: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). When the values of these metrics are smaller, the predictive model is more effective, indicating that the model is better at prediction. By evaluating these indicators, the predictive ability and performance of the model can be judged to be better or worse. The expressions are as follows:

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{9}$$

where y_i and $\hat{y}_i$ represent the actual and predicted data values of the i th predicted point, respectively; n denotes the total number of predicted data points.

II. C.Experimental design and data collection

In order to make the distribution of the collected experimental data more comprehensive and reasonable, a number of physiological parameters, including blood pressure, oxygen saturation, heart rate and temperature, were collected from 20 subjects in multiple states in the laboratory environment.

In order to validate the effectiveness of the Informer-based time series prediction model for elite athletes, the experimental environments of optimal athletic state, overtraining state, psychological stress state and environmental adaptation state were set up in the process of experimental data collection. Physiological parameters and exercise parameters were collected under the corresponding environments to complete the calibration of the experimental data. The settings of the experimental environments were as follows, and the data were collected for 20 minutes under different environments.

(1) Optimal competition state: virtual reality technology simulates the real competition scene, including audience noise, opponent's movement, venue lighting, etc.

(2) Over-training state: training in extreme environments such as low oxygen or high temperature.

(3) Psychological stress state: implanting high-pressure tasks in the training process.

(4) Environmental adaptation state: simulate the temperature and humidity of the competition scene and train in a similar material venue.

When the data collection of physiological parameters is carried out with the help of wireless health monitoring system, the sampling interval of body temperature, heart rate and blood oxygen saturation is set to 10s, which means that 6 groups of data can be collected per minute, and the sampling interval of blood pressure is 20s, which means that 3 groups of data can be collected per minute.

III. Analysis of fluctuations in physiological data and prediction of competitive status of elite athletes

III. A. Identification of physiological data for different athletic states

In order to validate the proposed model for predicting competitive state fluctuations in elite athletes, the physiological parameters of the experimental subjects were measured using a wireless health monitoring system. In this section, a total of 4 features of blood pressure, blood oxygen, heart rate, and body temperature were used for the identification of the experimental subjects' athletic state prediction. The physiological parameters of 20 experimental subjects were collected in four experimental environments, and the data category labels were obtained according to the experimental environments, and the amount of data of each of the four different categories was selected in 100 groups, respectively.The visualization of the physiological parameters of the 20 experimental subjects in different states is shown in Fig. 4.

As can be seen from the figure, the physiological parameters of the experimental subjects collected in different states have more obvious specificity. The body temperature of the athletes in the optimal competitive state was maintained between 36.5°C and 38°C, while the body temperature of the athletes in the overtraining state exceeded 39°C, and the athletes' temperatures were maintained below 39°C in the other two states. In terms of blood pressure measurements, athletes in the optimal competitive state had lower blood pressure than those in the remaining three states, and competitive intensity or psychological stress increased athletes' blood pressure indicators. In addition, athletes' heart rate changes had the same trend as blood pressure changes, and athletes in the optimal competitive state had a lower heart rate. For blood oxygen indicators, blood oxygen levels ranged from 86% to 96% in athletes in the overtrained state and from 92% to 99% in athletes in the environmentally adapted state. The measured psychometric data were used in the following prediction of athletes' competitive state prediction experiment.

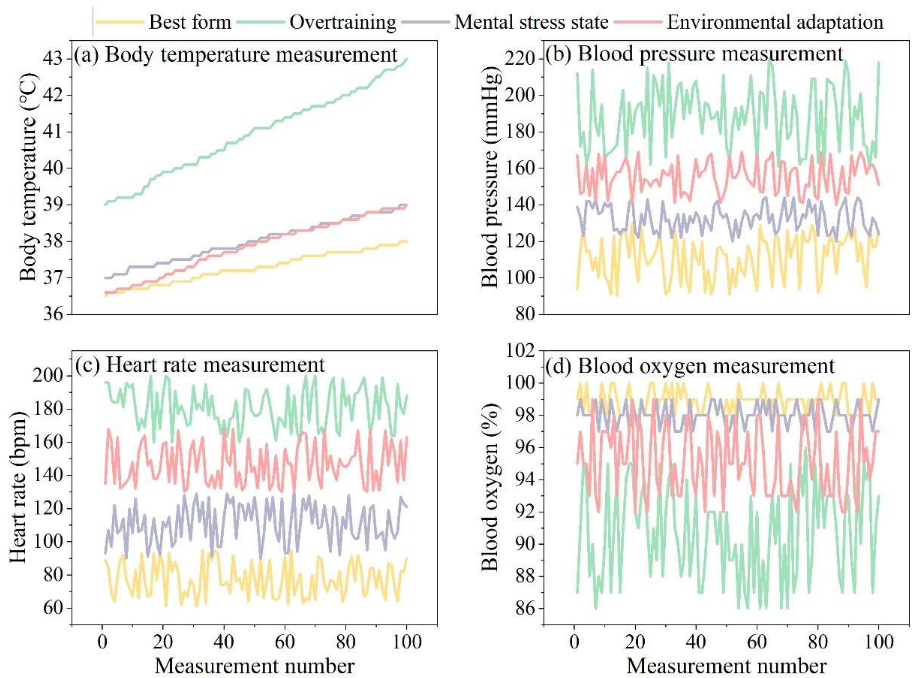

Figure 4: Visualization of physiological parameters in different states

III. B. Performance Analysis of Athletic Status Prediction for Athletes

In order to validate this paper's sports big data-driven prediction model of elite athletes' competitive status combined with time series analysis, this section does the corresponding prediction validation experiments on the dataset collected above. In addition to the 3 error metrics mentioned above, 2 metrics are introduced in this section. Namely, precision and recall. Precision is used to measure the accuracy of the result set, while recall is used to measure the security of the result set. Also, the definition of F1 value is given to examine the combined effect of the model. These 3 metrics are defined as follows:

$$\begin{aligned} \Pr ecision &= \frac{TP}{TP+FP} \\ \operatorname{Re} call &= \frac{TP}{TP+FN} \\ F1 &= 2\times\frac{\Pr ecision\times\operatorname{Re} call}{\Pr ecision+\operatorname{Re} call} \end{aligned} \tag{10}$$

TP and TN denote the number of correctly predicted positive and negative samples, respectively; FP and FN denote the number of incorrectly predicted positive and negative samples, respectively.

First, the accuracy under the energy consumption level in different competitive states of the athletes was predicted separately using the proposed model. The 10-fold validation method was used throughout the experiment. The dataset was divided into 100 copies, of which 80 were used to train the machine learning model and the remaining 20 were used as a test set. Table 1 shows the prediction results of the model athletic state in this paper. It can be observed that the model achieves the best results in predicting the environmental adaptation state with an F1 value of 0.962. Meanwhile, the model is able to achieve a better recall in predicting the optimal athletic state with 0.973.

Table 1: The prediction effect of model competition in this paper

Competitive state	Precision	Recall	F1
Best form	0.945	0.973	0.959
Overtraining	0.955	0.954	0.955
Mental stress state	0.943	0.942	0.943
Environmental adaptation	0.97	0.955	0.962

At the same time also compared the effect of different algorithms in different athletic status prediction of athletes, and the comparison algorithms include Plain Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). The evaluation metrics are the three error metrics described above, i.e., RMSE, MAE, and MAPE, and the results of the athletes' athletic status prediction errors of different algorithms are shown in Figure 5. From the figure, it can be seen that in predicting different athletic states of athletes, the prediction errors of this paper's algorithm are the lowest, and the best results are achieved. For example, in the best competitive state, the RMSE, MAE and MAPE predicted by this paper's algorithm are 0.058, 0.132 and 0.069 respectively, and the prediction results of the comparison algorithm are between 0.15 and 0.3. The results show that the Informer-based time series prediction model in this paper can achieve better prediction results of athletic status.

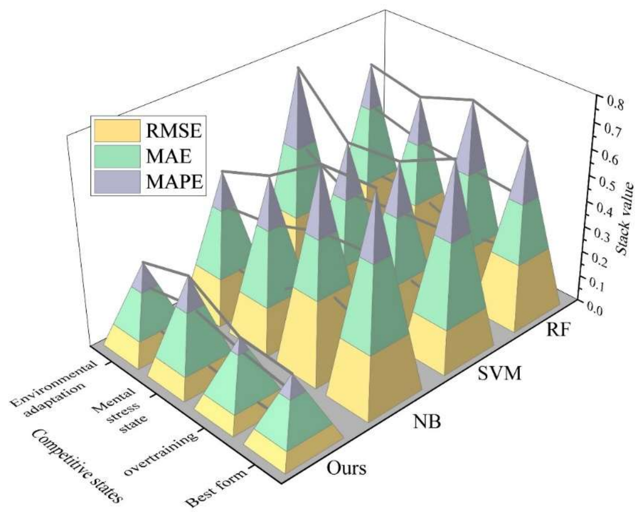

Figure 5: The results of the prediction error of the athletes' competitive state

III. C. Example Validation of Athletic State Prediction

In order to test the learning effect of the Informer-based time series prediction model, this section applies the prediction model to predict the body temperature, blood pressure, heart rate, and blood oxygen of 30 elite athletes in a sports academy. Because the physiological data of the athletes have a high correlation with their competitive status, the athletes' competitive status can be analyzed through the physiological prediction data.

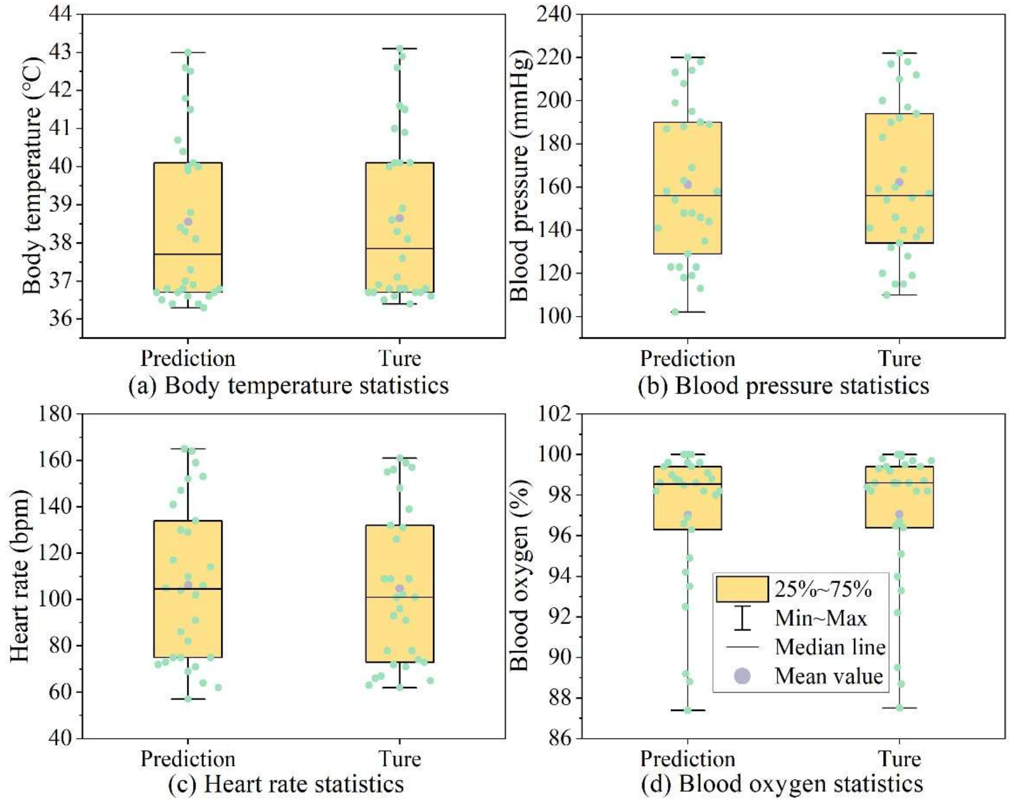

Figure 6: The predictive value of physiological data is compared to the real value

Figure 6 gives the results of the comparison between the predicted and real values of the model's physiological data, and (a)~(d) represent the comparison results of body temperature, blood pressure, heart rate and blood oxygen in turn. The data in the figure show that the model of this paper has high prediction accuracy, and the predicted values of body temperature, blood pressure, heart rate and blood oxygen for 30 elite athletes are basically consistent with the real values. Specifically, this paper predicts the body temperature of the 30 athletes to be 36.3°C to 43°C, with an absolute error between 0 and 0.5°C from the true value. In terms of blood pressure, the difference between the predicted value and the true value is between 0~9mmHg, which provides reference data for accurately predicting athletes' competitive status. Heart rate is one of the important indicators for predicting athletes' competitive state, and the increase of competitive intensity can significantly improve athletes' heart rate state. Among the 30 athletes studied, their heart rates were distributed between 60 and 165 under different competitive states. The difference between the predicted and true values of the model in this paper ranged from 0 to 9 bpm. Blood oxygen saturation is generally considered that it should not be lower than 94%, and in different competitive states, the blood oxygen saturation of athletes will be reduced accordingly. From Fig. (d), it can be seen that for the low-saturation athletes, this paper's model can also accurately identify their oxygen saturation, and the prediction error is between 0~0.4%. It can be seen that the use of this paper's prediction model has a good application for the prediction of physiological data of athletes' competitive state, and the competitive state behind the results of physiological data will be analyzed in the following.

In order to validate the Informer-based time series prediction model, the athletic status of 30 athletes was predicted with respect to the above mentioned athletes. The results of athletes' competitive state prediction are shown in Fig. 7. The model's competitive state prediction results were obtained by integrating the physiological data above and normalizing the combined data to obtain the predicted competitive state data distributed between [0,1]. Among them, [0,0.25) belongs to the optimal competitive state, [0.25,0.5) belongs to the psychological stress state, [0.5,0.75) belongs to the environmental adaptation state, and [0.75,1) belongs to the overtraining state. The prediction results of athletes' competitive state in the figure are basically consistent with the actual results of athletes' competitive state survey, showing the predictive validity of this paper's model, which can be applied to future competitions.

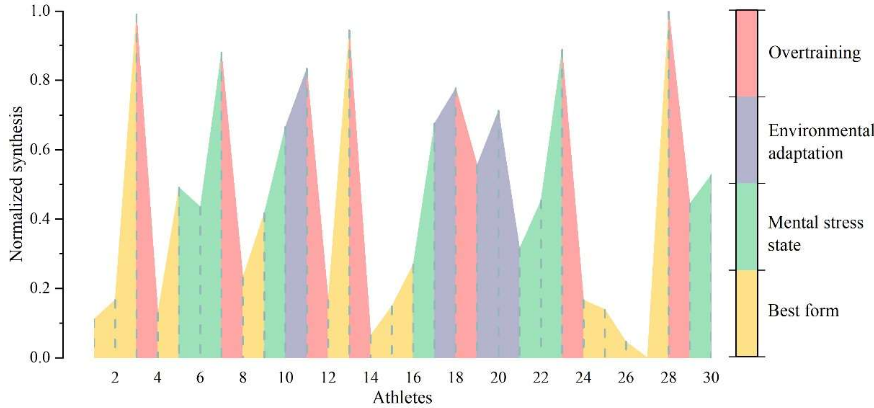

Figure 7: Athletes' competitive status prediction results

IV. Conclusion

The study simulated different competitive environments to make elite athletes in different competitive states, from which physiological data were collected under the states of optimal competitive state, overtraining state and so on. At the same time, a prediction framework for elite athletes' competitive state was formed by utilizing an Informer-based time series model. The framework can mine the potential relationship between the athlete's physiological data and the competitive state, and realize the prediction of the athlete's competitive state through deep learning data features.

In the overtraining state, the athlete's body temperature will rise sharply, exceeding 39℃, and the blood pressure and heart rate will also show an increasing trend, while the blood oxygen saturation will be lower than the normal value. The Informer-based prediction model had the highest recall in the optimal competitive state, and the RMSE, MAE, and MAPE in this state were 0.058, 0.132, and 0.069, respectively, which were lower than the comparison model. The prediction error of physiological data of this model is in the acceptable range and can accurately predict the competitive state of elite athletes.

This study not only extends the application scenarios of time series analysis in the field of sports, but also provides new ideas for the scientific and intelligent management of athletes' competitive status. In the future, more advanced deep learning technologies can be integrated with individual differences of athletes to develop training personalized optimization models that are adaptive to the characteristics of different athletes.

Funding

This research was supported by the North China Electric Power University, Tracking and Discrimination of Football Players by Integrating the Improved YOLOv5 and the Deep DIoU (Distance Intersection over Union) Strategy (No.: JLD898).

References

[1] Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. International Journal of Data Science and Analytics, 5(4), 213-222.
[2] Šuštaršič, A., Videmšek, M., Karpljuk, D., Miloloža, I., & Meško, M. (2022). Big data in sports: A bibliometric and topic study. Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy, 13(1), 19-34.
[3] Kaur, A., Kaur, R., & Jagdev, G. (2021). Analyzing and exploring the impact of big data analytics in sports sector. SN Computer Science, 2(3), 184.
[4] Herberger, T. A., & Litke, C. (2021). The impact of big data and sports analytics on professional football: A systematic literature review. Digitalization, digital transformation and sustainability in the global economy: risks and opportunities, 147-171.
[5] Jessop, A., & Baker III, T. A. (2019). Big data bust: Evaluating the risks of tracking NCAA athletes' biometric data. Tex. Rev. Ent. & Sports L., 20, 81.
[6] Mataruna-Dos-Santos, L. J., Faccia, A., Helú, H. M., & Khan, M. S. (2020, May). Big data analyses and new technology applications in sport management, an overview. In Proceedings of the 2020 international conference on big data in management (pp. 17-22).
[7] Kang, S. (2019). Utilization and Prospect of Big Data Analysis of Sports Contents. Convergence Security Journal, 19(1), 121-126.
[8] Kim, J. Y., Kim, J. H., Kang, E. W., Chang, T. I., Lee, Y. K., Park, K. S., ... & Park, Y. H. (2024). The Prediction of Dry Weight for Chronic Hemodialysis Athletes Using a Machine Learning Approach: Sports Health Implications. Journal of Sport Psychology/Revista de Psicología del Deporte, 33(1).
[9] Eid, A. I. A., Miled, A. B., Fatnassi, A., Nawaz, M. A., Mahmoud, A. F., Abdalla, F. A., ... & Mohamed, I. B. (2024). Sports prediction model through cloud computing and big data based on artificial intelligence method. Journal of Intelligent Learning Systems and Applications, 16(2), 53-79.

[10] Miah, J., Mamun, M., Rahman, M. M., Mahmud, M. I., Ahmad, S., & Nasir, M. H. B. (2022, December). Mhfit: Mobile health data for predicting athletics fitness using machine learning models. In 2022 2nd International seminar on machine learning, Optimization, and Data Science (ISMODE) (pp. 584-589). IEEE.
[11] Sweeting, A. J., Aughey, R. J., Cormack, S. J., & Morgan, S. (2017). Discovering frequently recurring movement sequences in team-sport athlete spatiotemporal data. Journal of Sports Sciences, 35(24), 2439-2445.
[12] Zhang, L., & Liang, F. (2021). Monitoring and analysis of athletes' local body movement status based on BP neural network. Journal of intelligent & fuzzy systems, 40(2), 2325-2335.
[13] Su, Z. (2022). Artificial intelligence in the auxiliary guidance function of athletes' movement standard training in physical education. Journal of circuits, systems and computers, 31(11), 2240001.
[14] Zhu, D., Zhang, H., Sun, Y., & Qi, H. (2021). Injury risk prediction of aerobics athletes based on big data and computer vision. Scientific Programming, 2021(1), 5526971.
[15] Baerg, A. (2017). Big data, sport, and the digital divide: Theorizing how athletes might respond to big data monitoring. Journal of Sport and Social Issues, 41(1), 3-20.
[16] Watanabe, N. M., Shapiro, S., & Drayer, J. (2021). Big data and analytics in sport management. Journal of Sport Management, 35(3), 197-202.
[17] Exel, J., & Dabnichki, P. (2024). Precision sports science: what is next for data analytics for athlete performance and well-being optimization?. Applied Sciences, 14(8), 3361.
[18] Wang, Y. (2021). Real-Time Collection Method of Athletes' Abnormal Training Data Based on Machine Learning. Mobile Information Systems, 2021(1), 9938605.
[19] Tan, X. (2023). Enhanced sports predictions: a comprehensive analysis of the role and performance of predictive analytics in the sports sector. Wireless Personal Communications, 132(3), 1613-1636.
[20] Feng, X., & Liu, X. (2025). Use predictive analytics and big data technologies to enhance badminton game strategy development and performance indicators. Journal of Computational Methods in Sciences and Engineering, 14727978251324125.
[21] Lu, X., Li, M., & Ye, C. (2024, February). Research on Sports Competition Result Prediction System Based on Computer Big Data Machine Learning Model. In 2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA) (pp. 1626-1630). IEEE.
[22] Zhu, H. (2022). Research on intelligent analysis strategies to improve athletes' psychological experience in the era of artificial intelligence. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 119, 110597.
[23] Chao, Z., Yi, L., Min, L., & Long, Y. Y. (2024). IoT-Enabled Prediction Model for Health Monitoring of College Students in Sports Using Big Data Analytics and Convolutional Neural Network. Mobile Networks and Applications, 1-18.
[24] Syed, L., Jabeen, S., Manimala, S., & Elsayed, H. A. (2019). Data science algorithms and techniques for smart healthcare using IoT and big data analytics. Smart techniques for a smarter planet: towards smarter algorithms, 211-241.
[25] Pastrana, J. L., Reigal, R. E., Morales-Sánchez, V., Morillo-Baro, J. P., Mier, R. J. R. D., Alves, J., & Hernández-Mendo, A. (2019). Data mining in the mixed methods: application to the study of the psychological profiles of athletes. Frontiers in Psychology, 10, 2675.
[26] Ma, W., & Guo, B. (2024). Advanced neurological activity status of athletes based on big data technology. Heliyon, 10(17).
[27] Guo, X., Liu, P., & Li, T. (2024, September). Sports Injury Prediction and Prevention: Analysis Methods Based on Big Data and Artificial Intelligence. In 2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 60-63). IEEE.
[28] Yang, H., Zhang, S., Zhang, J., & Wang, C. (2024). Evaluating the performance of athletes in various sports using data mining and big data analytics. Soft Computing, 28(4), 2875-2890.
[29] Li, W. (2024). A Big Data Approach to Forecast Injuries in Professional Sports Using Support Vector Machine. Mobile Networks and Applications, 1-17.
[30] Suvarna, S., Sindhu, C., Nair, S., & Kolluru, A. N. (2023, February). Fitness Prediction in High-Endurance Athletes and Sports Players Using Supervised Learning. In International Conference On Innovative Computing And Communication (pp. 303-316). Singapore: Springer Nature Singapore.
[31] Song, J., & Zhong, J. (2022). Fuzzy Theory in the Prediction of Athletes' Competitive State Based on Information Security. Mobile Information Systems, 2022(1), 6554082.
[32] Bahman Rostami Tabar & Rob J. Hyndman. (2025). Hierarchical Time Series Forecasting in Emergency Medical Services. Journal of Service Research,28(2),278-295.
[33] Zhihe Yang,Jiandun Li,Haitao Wang & Chang Liu. (2025). An Informer Model for Very Short-Term Power Load Forecasting. Energies,18(5),1150-1150.
[34] Abhinav Suresh,Henning Schlömer,Baran Hashemi & Annabelle Bohrdt. (2025). Interpretable correlator Transformer for image-like quantum matter data. Machine Learning: Science and Technology,6(2),025006-025006.
[35] Klavdia Zemlianova,Amitabha Bose & John Rinzel. (2025). Author Correction: Dynamical mechanisms of how an RNN keeps a beat, uncovered with a low-dimensional reduced model. Scientific Reports,15(1),12256-12256.