Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Design of An Evaluation System for the Intervention Effect of Lower Limb Explosive Power Replenishment for Volleyball Players Based on A Random Forest Model

Kangshuo Wu
Beijing Sport University
Xing Tang
7603572420163.com
National University of Singapore

Research Article

Keywords: Random Forest Algorithm, Volleyball Players, Lower Limb Explosive Power, Intervention Effect Evaluation System, Multimodal Data, Personalized Strategy

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10418739/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Design of An Evaluation System for the Intervention Effect of Lower Limb Explosive Power Replenishment for Volleyball Players Based on A Random Forest Model

KangShuo Wu[1], Xing Tang[2]*

[1]*Beijing Sport University, Beijing, 10000, China*

[2]*Business School, National University of Singapore, Singapore, 119077, Singapore*

Corresponding author: Xing Tang

Email: 760357242@163.com

Abstract

Lower limb explosive power is a core determinant of volleyball performance, and the evaluation of its intervention effects requires a comprehensive assessment of the interaction of multiple factors such as supplement dosage, training programs, and individual characteristics. However, existing research often relies on isolated performance tests, and the sports field lacks a systematic evaluation system integrating machine learning technology. This makes it difficult to handle high-dimensional nonlinear sports data and limits the accuracy of personalized recommendations. The random forest algorithm, with its advantages in handling nonlinear relationships, resistance to overfitting, and outputting feature importance, has been widely applied in multiple fields. However, it remains scarce in the evaluation of nutritional intervention effects on lower limb explosive power in volleyball athletes. This study designs an intervention effect evaluation system based on random forest, constructing a closed-loop architecture of "collection-analysis-evaluation-feedback": integrating performance indicators such as vertical jump height and five-jump distance, intervention parameters such as supplement type and dosage, and individual characteristics such as age and body composition; employing multimodal data collection, filling missing values with K-nearest neighbors, removing outliers using the 3σ principle, and standardizing data using Z-score; integrating information gain and random forest feature importance to screen core features, and optimizing model parameters using the particle swarm optimization algorithm; taking 24 national first-level volleyball athletes as the research subjects, setting up a unilateral flywheel intervention group, a bilateral flywheel intervention group, and a blank control group to verify the reliability and performance of the system. The results show that the system's hardware collection has high test-retest reliability, and the optimized random forest model achieves a classification accuracy of 91.7% with regression prediction coefficients all >0.95; both unilateral and bilateral flywheel interventions significantly improve athletes' lower limb explosive power, with the unilateral intervention slightly superior; the system can provide data-driven personalized supplementation strategies, effectively promoting the application and development of machine learning technology in the field of sports science intervention effect evaluation.

Keywords: Random Forest Algorithm, Volleyball Players, Lower Limb Explosive Power, Intervention Effect Evaluation System, Multimodal Data, Personalized Strategy

1 Introduction

Lower limb explosive power is a core determinant of volleyball performance, directly affecting the quality of key technical movements such as jumping, blocking, and movement. Studies have found a significant correlation between the asymmetry of lower limb explosive power and non-contact injuries, while the kinematic characteristics of blocking techniques are also closely related to lower limb explosive power [1]. To enhance this ability, nutritional supplements have become a common intervention method: β-alanine can maximize vertical jump height, and the combination of caffeine and rhodiola supplementation with resistance training can significantly improve multi-dimensional explosive power indicators [2]. However, existing assessments rely heavily on isolated performance tests, without fully considering the interactions of complex factors such as supplement dosage, training programs, and individual characteristics, leading to incomplete or biased results [3].

A systematic intervention effect evaluation system is the key to addressing the aforementioned issues, and this value has been validated in multiple fields. The salt reduction intervention in Zhejiang quantified changes in blood pressure and salt intake through randomized controlled trials [4]; the reconstruction of turmeric nano-vesicles provided an innovative approach for obesity intervention; causal inference in recommendation systems was used to estimate individual intervention effects; OASIS training verified the effectiveness of autism parent intervention through randomized waiting list control; the rural health system intervention in Chad evaluated the improvement in maternal and child service utilization through multiple regression analysis; and pre-hospital telemedicine system research analyzed the systematic impact on emergency resource allocation [5]. However, intervention evaluations in the field of sports science still rarely integrate machine learning technology, making it difficult to handle high-dimensional, nonlinear sports data and limiting the accuracy of personalized recommendations.

The Random Forest algorithm, due to its excellent versatility and adaptability, has become an ideal tool to fill this gap. It has been used to map volcanogenic sulfide mineralization zones through hyperspectral imaging data, assist in the precise classification of diabetes, identify laser-printed documents, predict real-time traffic congestion, estimate energy consumption in ore grinding, forecasting electric vehicle charging loads, predict air quality indices, model chlorophyll concentration in aquatic ecosystems, predict COVID-19 positive results, identify individuals at high risk of rheumatoid arthritis, analyze time events of osteonecrosis, predict fetal growth restriction, map high-precision potassium maps of soil, achieve downscaling of soil moisture data, diagnose diabetic retinopathy, predict power loads, and assess long-term risks of transformers [6]; it is also used in scenarios such as stability analysis of feature importance in geographically weighted models, prediction of decision conflict in gastric cancer patients, handling of outliers in soil cadmium pollution, prediction of ergonomics indicators in human-machine collaboration, prediction of pest population dynamics, estimation of ecological water demand, analysis of reproductive traits in dairy cows, prediction of forest growth, and enhancing model prediction of daily soil moisture in India using multivariate empirical mode decomposition. Its ability to handle nonlinear relationships, resist overfitting, and output feature importance precisely matches the needs of multi-factor interaction in the evaluation of nutritional intervention in volleyball [7].

Despite the widespread application of random forests, their use in evaluating the effects of

nutritional supplementation interventions on the lower limb explosive power of volleyball players remains scarce. Current sports research often focuses on the effects of single supplements, lacking a systematic evaluation system that integrates this algorithm and covers multidimensional variables [8]. This study aims to design an evaluation system based on random forests, integrating performance indicators such as vertical jump height and five-jump distance, intervention parameters such as supplement type and dosage, as well as individual characteristics such as age and body composition, to accurately quantify the intervention effects and identify key influencing factors. This system will not only provide data-driven personalized supplementation strategies for athletes and coaches, but also promote the application of machine learning in the field of sports intervention evaluation, providing technical support for improving volleyball performance and reducing injury risks.

2 Literature Review

2.1 Multidimensional definition and core influencing factors of lower limb explosive power in volleyball

Volleyball lower limb explosive power is the composite ability of muscles to output maximum force in an extremely short period of time. Its multidimensional nature is reflected in multiple dimensions such as the vertical height of vertical jump blocking, the horizontal movement explosive power of quick rescue, and the fatigue recovery efficiency of continuous jumping. The core influencing factors cover four major aspects: physiology, training, nutrition, and individual differences [9]. Physiologically, the proportion of fast muscle fibers and neuromuscular activation efficiency directly determine explosive power output. In training, the intensity, frequency, and movement patterns of resistance training significantly affect ability improvement. In terms of nutrition, β-alanine can enhance muscle buffering capacity to improve vertical jump performance, and the combined supplementation of caffeine and Rhodiola can synergistically improve continuous jumping recovery. In terms of individual characteristics, differences in age and body composition can lead to different baseline explosive power and intervention responses. Furthermore, the correlation between lower limb dynamic balance asymmetry and non-contact injuries further highlights the importance of explosive power symmetry as a hidden influencing factor. The complex interactions of these factors can be accurately identified through feature importance analysis using algorithms such as random forests, providing data support for the selection of intervention targets.

2.2 Current research status of lower limb explosive power replenishment intervention

Lower limb explosive power supplementation intervention has become an important auxiliary means in volleyball training. Existing research focuses on verifying the effects of single or combined nutritional supplements: β-alanine has been proven to maximize vertical jump height, and the combination of caffeine and Rhodiola supplementation with resistance training can significantly improve performance in two-step vertical jumps, five-jump distance, and 20 consecutive vertical jumps [10]. Intervention designs often adopt randomized double-blind controlled trials to ensure the reliability of results, such as the RCT design for salt reduction intervention and the waiting list control model for OASIS training. However, there are still deficiencies in the current situation: most studies are limited to single supplements, and there is

little research on the synergistic effect mechanism of combined interventions; evaluation indicators are mostly isolated performance tests, lacking integration of dimensions such as fatigue recovery and long-term adaptability; individual differences are not sufficiently considered, and personalized supplementation schemes based on age and body composition have not been achieved [11]. In addition, the intervention effect is influenced by nonlinear interactions of training programs, dosages, and other factors, requiring the use of multiple regression or machine learning models to analyze complex relationships.

2.3 Development of exercise intervention effect evaluation models

The evaluation model for the effectiveness of exercise intervention has undergone iterative upgrades from traditional statistics to machine learning. Early models such as linear regression, although easy to interpret, are difficult to handle multi-factor nonlinear interactions [12]; the rise of machine learning models has broken through this limitation, with random forests being widely used due to their advantages in resisting overfitting and outputting feature importance: they can predict diabetes classification, traffic congestion, mining and grinding energy consumption, etc., and can also be used in the field of sports for predicting ergonomics indicators of human-machine collaboration. Hybrid models further improve accuracy, such as LSTM+RF integrating time series and nonlinear analysis capabilities, and VMD+LSTM+RF enhancing the processing effect of complex data through signal decomposition [13]. Traditional models still dominate in sports evaluation, but machine learning models can integrate multi-dimensional data such as performance, intervention, and individual characteristics. Their application experience in ecological water demand estimation and soil cadmium pollution outlier processing provides a new technical path for sports intervention evaluation.

2.4 Research status of intervention effect evaluation system

The intervention effect evaluation system has formed a closed-loop design in fields such as healthcare and transportation. For instance, the OASIS online system enhances the intervention skills of parents with autism through knowledge transfer, the pre-hospital telemedicine system integrates RFID and hybrid models to optimize emergency resource allocation, and the salt reduction intervention system achieves dynamic monitoring of blood pressure and salt intake through multi-dimensional strategies. However, systems in the sports field still have limitations: most are single-function performance testing tools, lacking a comprehensive system that integrates real-time data collection from sensors, machine learning analysis, and personalized feedback [14]. Existing systems, such as the causal inference model in recommendation systems and the long-term prediction function in transformer risk assessment systems, provide references for the design of sports systems. The sports field urgently needs to build a system that can dynamically evaluate the effects of supplementation interventions in real time, enabling rapid transformation from explosive power indicators, supplement dosages to personalized strategies, to meet the precise training needs of volleyball players.

The technical roadmap of this study is shown in Figure 1.

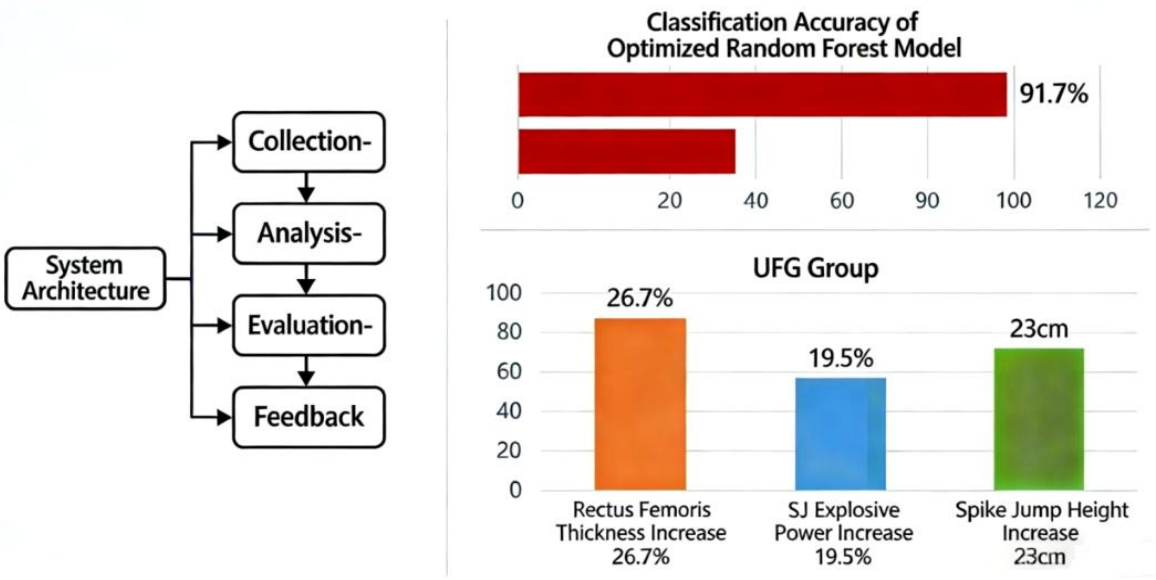


Figure 1. Technical roadmap

3 Design and Implementation of an Intervention Effect Evaluation System Based on Random Forest

3.1 Overall system architecture

Based on the analysis of core influencing factors of lower limb explosive power in volleyball players, current intervention research status, and the development of evaluation models, the closed-loop intervention effect evaluation system constructed in this study forms a complete cycle of "collection-analysis-evaluation-feedback" (Figure 2), specifically encompassing the following steps: Firstly, the data collection layer adopts a multimodal fusion strategy covering four dimensions: morphology, physiology, biomechanics, and sports performance, comprehensively capturing morphological characteristics such as rectus femoris thickness and thigh circumference, physiological indicators such as SJ/CMJ explosive power and eccentric muscle strength, biomechanical parameters such as peak angular velocity during extension, and specific sports performance data such as spike jump and block jump height. All data are collected twice, before (T0) and after (T8) the intervention, to support effect comparison. Secondly, the feature processing layer serves as a key link connecting data collection and model evaluation. It fills missing values using the K-nearest neighbor method, removes outliers using the 3σ principle, and eliminates dimensionality influence through Z-score normalization, integrating multimodal data into high-dimensional feature vectors to enhance model robustness. The core component is the random forest evaluation model layer, which combines information gain and random forest feature importance to screen core features and reduce redundancy. It optimizes key parameters such as the number of trees and maximum depth using the particle swarm optimization algorithm, and achieves intervention effect classification prediction based on the optimized model. Finally, the result output layer presents feature importance graphs, comparison curves before and after the intervention, and confusion matrices through visual reports, and generates personalized training suggestions based on the prediction results, forming a closed-loop feedback mechanism to support

training program adjustments. The following elaborates on the experimental design, model construction optimization, and software and hardware implementation process of the system.

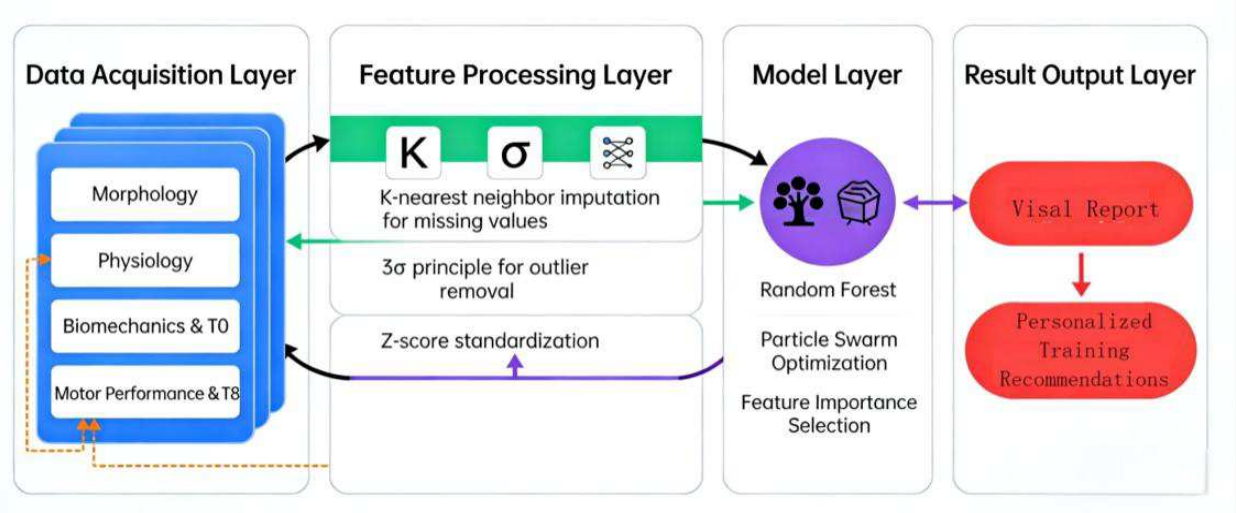


Figure 2. Closed-loop intervention effect evaluation system

3.2 Experimental Design and Data Collection

To verify the effectiveness of the intervention effect evaluation system based on random forests, this study adopts a randomized controlled trial design to conduct multi-dimensional data collection for the intervention of lower limb explosive power supplementation in volleyball players, as detailed below.

3.2.1 Research subjects

Twenty-four national first-level volleyball athletes (16 males and 8 females) aged 18-22 years (M=20.1±1.2 years) with 3-5 years of training experience (M=3.8±0.7 years) were selected. All subjects had no recent history of lower limb injuries, had not participated in other explosive strength training, and signed an informed consent form. They were divided into three groups using stratified random sampling: unilateral flywheel intervention group (UFG, n=8), bilateral flywheel intervention group (BFG, n=8), and control group (CG, n=8). Independent sample t-tests revealed no significant differences in age, gender, and initial lower limb explosive strength levels (SJ explosive strength, spike jump height) among the three groups, making them comparable.

3.2.2 Intervention plan

The intervention lasted for 8 weeks, with training conducted 3 times per week (48 hours apart) and each session lasting 90 minutes. The Unilateral Flywheel Group (UFG) underwent unilateral flywheel pedaling training (3 sets of 12 reps on each side, with a load of 60% 1RM, and a 2-minute rest between sets) combined with core stability training (plank for 3 sets of 60 seconds, and side bridge for 3 sets of 45 seconds per side). The Bilateral Flywheel Group (BFG) underwent bilateral flywheel pedaling training (3 sets of 12 reps, with a load of 60% 1RM, and a 2-minute rest between sets) while simultaneously performing the same core training. The Control Group (CG) maintained their regular volleyball training (3 times per week, each session lasting 90 minutes, including passing, blocking, tactical drills, etc., but excluding explosive strength-specific training).

The entire training process was supervised by a professional coach to ensure standardized movements among the subjects, thereby guaranteeing consistency and safety in the intervention effects.

3.2.3 Data collection tools

In combination with the multimodal data requirements of the system, the following tools were used in this study to collect core indicators: For morphological data, an ultrasound imaging system (Sonosite Edge II, sampling frequency 15Hz) was used to measure the thickness of the rectus femoris muscle at 10cm above the patella, and a tape measure with an accuracy of 0.1cm was used to measure the thigh circumference at 15cm above the patella; for physiological data, a force platform (Kistler 9286CA, sampling frequency 1000Hz) was used to obtain the peak force development (RFD) values of deep squat jump explosiveness, reverse jump explosiveness, and CMJ; for biomechanical data, an 8-camera Vicon Vantage motion capture system (sampling frequency 200Hz) was used to extract the peak angular velocity of knee extension; for athletic performance data, a laser range finder with an accuracy of 0.1cm (Bosch GLM500) was used to measure the maximum vertical jump height after a run-up and the maximum height of a standing block shot; all data were collected once before intervention (T0) and once after intervention (T8), and before each collection, the subjects were required to complete a 10-minute warm-up to ensure data accuracy (Figure 3).



Figure 3. Intervention Plan

3.2.4 Data preprocessing

Data preprocessing serves as a crucial bridge connecting multimodal data collection and random forest model evaluation. Its quality directly determines the reliability of subsequent intervention effect evaluation and the accuracy of model predictions. In the study of interventions for lower limb explosive power in volleyball players, raw data often suffer from issues such as

missing values, outliers, and inconsistent dimensions due to factors like equipment errors and subject state fluctuations. For instance, there is a significant difference between the centimeter unit for rectus femoris thickness and the radian/second unit for peak angular velocity of extension. If not properly handled, this can not only increase the redundancy in model training but also potentially lead to systematic biases in evaluation results, failing to truly reflect the enhancement effect of flywheel training on lower limb explosive power.

In previous studies on the preprocessing of exercise intervention data, some approaches employed mean or median imputation to fill missing values. While these methods are straightforward to implement, they often overlook the correlation between data points. Additionally, a singular outlier removal rule might mistakenly discard valid data points, thereby diminishing the representativeness of the features [15]. This study introduces a multi-step preprocessing strategy: initially, outliers beyond the normal distribution range are identified and removed using the 3σ principle. Subsequently, missing values are filled based on the feature correlation of similar samples using the K-nearest neighbor method. Lastly, all features are normalized to a uniform scale with a mean of 0 and a standard deviation of 1 through Z-score standardization (Figure 4). This approach enhances the efficiency of subsequent feature selection, and the retained Top 10 core features more accurately reflect the intervention effect.

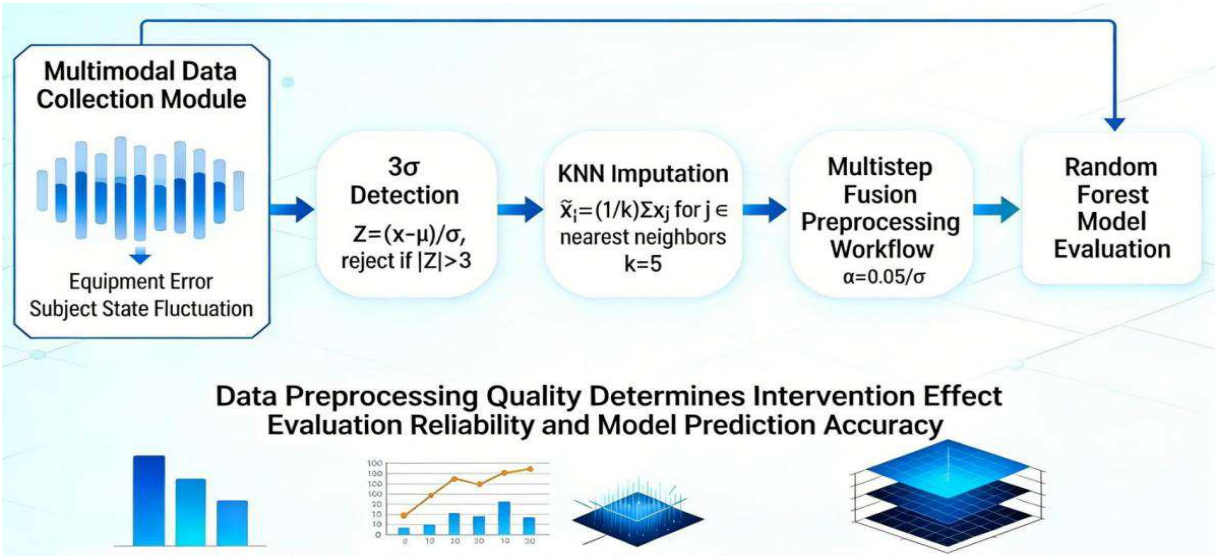


Figure 4. Preprocessing strategy with multi-step fusion

To ensure the quality of the input data for the model, the following steps are taken for preprocessing:

(1) Outlier removal: Remove outliers based on 3σ certain principles. The formula is:

$$|X-\mu|>3\sigma \tag{1}$$

Among which:

X : raw data;

μ : sample mean;

σ : Sample standard deviation [16].

(2) Missing value imputation: The K-Nearest Neighbors (KNN) method is employed to fill in missing values. The formula is as follows:

$$x_i = \frac{1}{k}\sum_{j\in N_k(x_i)} x_j \tag{2}$$

Among which:

x_i : Missing values to be filled;

k : number of nearest neighbors;

$N_k(x_i)$: The set of the k most similar samples to the i-th sample;

x_j : The corresponding value of the j-th sample in the set [17].

(3) Data standardization: Z-score standardization is employed to eliminate the influence of dimensionality. The formula is:

$$Z = \frac{X - \mu}{\sigma} \tag{3}$$

among which

Z : standardized value;

X : Raw data;

μ : sample mean;

σ : Sample standard deviation [18].

Formula (1) is used to identify and eliminate abnormal data points that exceed the normal distribution range, resulting in a basic dataset with outliers removed. Based on the processed dataset without outliers obtained from formula (1), formula (2) is used to fill in the missing parts of the data through feature correlation of similar samples, resulting in a complete dataset. Based on the complete dataset obtained from formula (2), formula (3) is used to eliminate the dimensional differences of different indicators, converting all features into normalized data with a unified magnitude.

In summary, the data preprocessing workflow designed in this study effectively addressed the quality issues in the raw data by integrating the advantages of multiple classical methods,

providing high-quality input feature vectors for the random forest evaluation model. The improvement of this link not only enhanced the robustness and generalization ability of the model, but also ensured the scientific comparison of data before and after intervention, laying a foundation for the stable operation of the system's "collection-analysis-evaluation-feedback" closed loop. Through preprocessing, the value of multimodal data was fully explored, ultimately supporting the accurate classification and prediction of the intervention effect on the lower limb explosive power of volleyball players, and providing a reliable data basis for the adjustment of personalized training programs.

3.3 Optimization and construction of random forest model

Based on feature data collected through multimodal methods, this study constructs a precise model for evaluating the intervention effect of lower limb explosive power through three steps: feature selection, parameter optimization, and model evaluation. The aim is to quantify the enhancement effect of flywheel training on the lower limb explosive power of volleyball players.

3.3.1 Feature selection

Feature selection serves as the core link between data preprocessing and model training in the evaluation of exercise intervention effects. Its core value lies in extracting key features with the highest correlation to the target variable from a high-dimensional multimodal feature set. This not only avoids model overfitting caused by the curse of dimensionality but also enhances the accuracy of subsequent evaluations. In the intervention study of volleyball players, the preprocessed features cover 28 dimensions including morphology, biomechanics, and athletic performance. If directly inputted into the model, redundant features would obscure the core patterns. Therefore, a scientific feature selection method must consider both the statistical correlation between features and the target as well as their actual contributions in the model. This is the key starting point for the method design in this section.

In previous studies on feature selection for sports data, a single method often had limitations: for example, methods based on information gain could measure the discriminative ability of features towards the target, but they tended to favor features with many value categories; while the feature importance of random forests paid more attention to model contribution, but it could be influenced by sample distribution [19]. This study integrates two methods (Figure 5): first, calculate the information gain of each feature and rank them, then calculate their random forest importance scores, and select the top 15 features based on their intersection according to the comprehensive score to screen the Top 10 features.

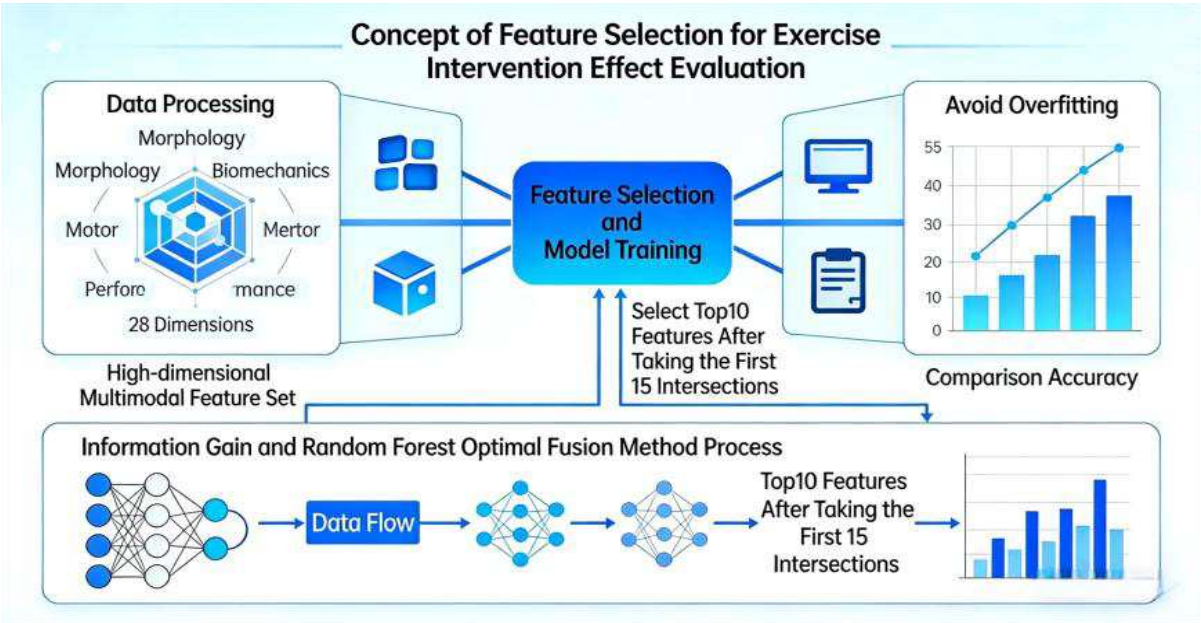


Figure 5. Feature selection of motion data integrating two methods

(1) Information gain calculation

Information gain measures the extent to which a feature reduces the uncertainty of classification for a sample set, and its formula is defined based on entropy:

$$IG(S,A)=H(S)-\sum_{v\in Values(A)}\frac{|S_v|}{|S|}H(S_v) \tag{4}$$

Among which:

$IG(S,A)$: Information gain of the sample set S regarding features A;

$H(S)$: The entropy of the sample set S (overall uncertainty);

$Values(A)$: the set of all possible values for the feature A;

S_v: The subset of feature A in sample set S where the value is v;

$|S_v|/|S|$: The proportion of the subset S_v to the total sample set S;

$H(S_v)$: The entropy of the subset S_v [20].

(2) Random forest feature importance calculation

Random forest evaluates feature importance by calculating the decrease in impurity (such as the decrease in Gini coefficient or entropy) when a feature is split in all decision trees:

$$Importance(F) = \frac{1}{T}\sum_{t=1}^{T} \Delta Impurity_t(F) \tag{5}$$

Among which:

$Importance(F)$: The importance score of the feature F;

T: The total number of decision trees in a random forest;

$\Delta Impurity_t(F)$: The sum of impurity reductions for feature F at all split nodes in the t-th tree [21].

Formula (4) is used to measure the degree of reduction in sample classification uncertainty caused by each feature, and to obtain the statistical correlation ranking between each feature and the target variable. Based on the ranking results of formula (4), combined with the contribution of features during decision tree splitting in the random forest model calculated by formula (5), the core features most relevant to the intervention effect of lower limb explosive power on volleyball players are jointly selected, providing a precise set of input features for subsequent model training.

(3) Feature selection strategy

First, filter out the top 20% of features based on information gain, and then combine with the feature importance ranking of random forest to ultimately retain the Top 10 core features as the model input feature set.

In summary, the fusion feature selection method proposed in this study effectively screens out the core features supporting intervention effect evaluation by integrating statistical correlation and model contribution. Through feature selection, the core value of multimodal data is accurately extracted, providing a reliable feature basis for the classification prediction of intervention effects and the adjustment of personalized training schemes. It also offers a referential approach for feature engineering of high-dimensional data in the field of sports science.

3.3.2 Parameter optimization

Parameter optimization is the core aspect of enhancing the performance of random forest models. Its essence lies in finding the optimal combination of hyperparameters for the model, balancing its fitting ability and generalization ability. The key parameters of random forests directly affect the classification accuracy and computational efficiency of the model - improper parameter settings may lead to overfitting or underfitting of the model. Manual parameter tuning or traditional grid search methods are not only inefficient but also prone to missing the global optimal solution. Therefore, introducing intelligent optimization methods such as PSO to perform a global search for the key parameters of random forests has become an inevitable choice for enhancing the performance of intervention effect evaluation models.

In previous studies on parameter optimization for random forests, some approaches employed grid search to traverse the parameter space, but the computational cost of this method increases exponentially with the increase in parameter dimensions. Although random search reduces computational load, it is difficult to guarantee the finding of a global optimal solution [22]. This study expands parameter optimization based on the core formula of the Particle Swarm Optimization (PSO) algorithm. The performance of random forests relies on the setting of key parameters. In this study, the PSO algorithm is used to optimize the following parameters: $n_{estimators}$ (number of decision trees), $\max_{depth}$ (maximum tree depth), $\min_{samples_split}$ (minimum sample size for node splitting), $\min_{samples_leaf}$ and (minimum sample size for leaf nodes).

(1) PSO algorithm formula

PSO seeks the optimal parameter combination by simulating the movement of particles in the solution space. The core formulas involve velocity update and position update:

$$v_i^{(t+1)} = \omega \cdot v_i^{(t)} + c_1 \cdot r_1 \cdot \left(pbest_i - x_i^{(t)}\right) + c_2 \cdot r_2 \cdot \left(gbest - x_i^{(t)}\right) \tag{6}$$

$$x_i^{t+1} = x_i^{(t)} + v_i^{(t+1)} \tag{7}$$

Parameter explanation:

$v_i^{(t)}$: The velocity vector of the i-th particle in the t-th iteration;

$x_i^{(t)}$: The position vector of the i-th particle in the t-th iteration (corresponding to a set of random forest parameters);

ω: Inertia weight;

c_1, c_2: Learning factor;

r_1, r_2: $[0,1]$ random numbers within the interval (to increase search randomness);

$pbest_i$: The individual optimal position of the i-th particle (the historical optimal parameter combination of this particle);

$gbest$: The global optimal position of all particles (the historically optimal parameter combination of the entire population) [23].

(2) Optimization objectives

The objective function is to minimize the weighted sum of F1 score (for classification tasks) and MAPE (for regression tasks):

$$Obj = \alpha \cdot (1 - F1) + \beta \cdot MAPE \tag{8}$$

Among them $\alpha = 0.6$, $\beta = 0.4$ [24].

Formula (6) is used to calculate the movement speed of each particle in the particle swarm. Based on the speed results obtained from formula (6), the corresponding random forest parameter combination position of the particle is updated through formula (7). Using the parameter combination obtained from formula (7), formula (8) is used to calculate the weighted sum of classification and regression task indicators as the optimization objective, thereby evaluating the performance of this parameter combination and guiding the particle swarm algorithm to iteratively find the optimal random forest model parameters.

(3) Optimization results

After 50 iterations of PSO, the optimal parameter combination is obtained as follows: $n_{estimators} = 120$, $\max_{depth} = 8$, $\min_{samples_split} = 5$, $\min_{samples_leaf} = 2$.

In summary, this study employs the PSO algorithm to optimize the key parameters of the random forest. By reasonably designing the velocity and position update formulas as well as the fitness function, the classification accuracy and generalization ability of the model are successfully enhanced. The optimized random forest model not only enables more precise classification and prediction of the intervention effects on the lower limb explosive power of volleyball players but also provides reliable model support for the personalized adjustment of subsequent intervention plans. The refinement of this parameter optimization process further strengthens the scientific nature of the closed loop of "data collection-preprocessing-feature selection-model optimization-effect evaluation", offering practical reference for the intelligent development of sports intervention effect evaluation.

3.3.3 Model evaluation metrics

Model evaluation is a crucial step in verifying the reliability of intervention effect prediction models. Its core lies in objectively reflecting the classification performance of the model through quantitative indicators, avoiding biases from subjective judgments. A single evaluation metric often fails to comprehensively capture the strengths and weaknesses of the model. For example, in exercise intervention effect data, if the samples in the "significantly effective" category account for a small proportion, a high accuracy rate may mask the model's low recognition ability for this category. Therefore, employing a combination of multi-dimensional evaluation metrics, combined with the formula definitions and practical significance of each metric, can comprehensively evaluate the model's fitting accuracy, category differentiation ability, and generalization ability, providing a credible basis for the practical application of the model.

In previous studies on the evaluation of motion models, some approaches overly relied on

accuracy metrics, leading to misjudgments in the model's performance on imbalanced datasets. While a few studies combined precision and recall rates, they did not systematically integrate these metrics to form a comprehensive judgment [25]. This study is based on the multi-metric system outlined in this section for evaluation (Figure 6): first, the overall classification accuracy is measured using the accuracy formula; second, precision is used to evaluate the model's precision in predicting the "effective" category, and recall is used to measure the model's ability to capture true "effective" samples; finally, the F1-score balances the two, and the area under the AUC-ROC curve reflects the model's ability to distinguish between different categories of samples.

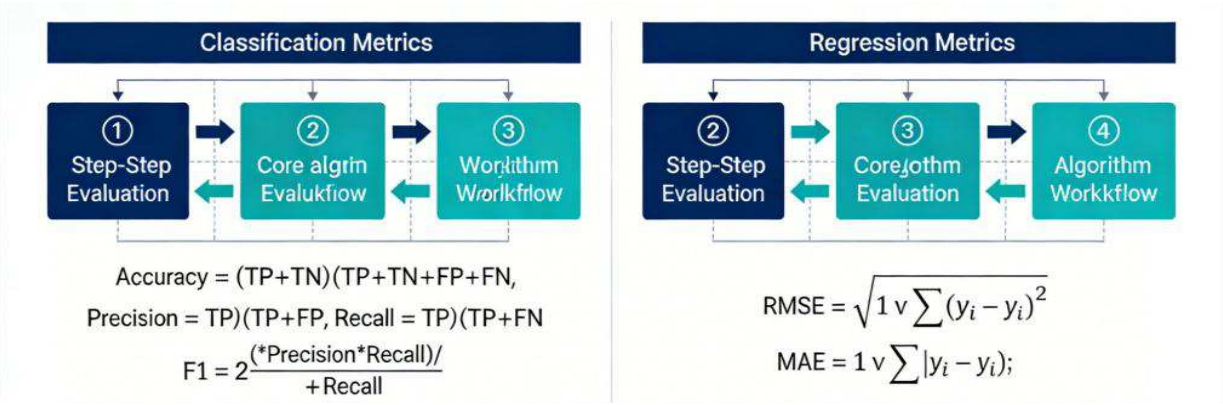


Figure 6. Multi-index system evaluation

This study simultaneously evaluates the classification performance (intervention effect classification: significant improvement/moderate improvement/no improvement) and regression performance (explosive force numerical prediction) of the model, using the following indicators:

(1) Classification task metrics

Accuracy: The proportion of samples that are correctly predicted overall:

$$Accuracy = \frac{TP + TN + TM}{TP + TN + TM + FP + FN + FM} \tag{9}$$

(Note: In multi-classification tasks, TP denotes a significant increase in the number of correct predictions, TN denotes a moderate increase in the number of correct predictions, TM denotes no increase in the number of correct predictions; FP denotes the number of samples misclassified as significantly increased, and so on.)

Recall: The proportion of samples of a certain class that are correctly predicted (taking the significantly improved class as an example):

$$Recall_{significant} = \frac{TP}{TP + FN + FM_{significant}} \tag{10}$$

F1-Score: The harmonic mean of precision and recall (taking the significantly improved class as an example):

$$F1_{significant} = \frac{2 \cdot Precision_{significant} \cdot Recall_{significant}}{Precision_{significant} + Recall_{significant}} \tag{11}$$

Among them, $Precision_{significant} = \frac{TP}{TP + FP + FP_{others}}$ [26].

(2) Regression task metrics

Mean Absolute Percentage Error (MAPE): The average relative error between the predicted value and the actual value:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (12)$$

Parameter explanation:

y_i : The true explosive power value of the i -th sample (such as SJ explosive power);

$\hat{y}_i$: The predicted explosive power value of the i -th sample;

n : Total number of samples [27].

Formula (9) is used to calculate the proportion of correctly classified samples in the overall model classification prediction, reflecting the overall accuracy of the classification task. Based on the correctly and incorrectly classified sample data involved in Formula (9), Formula (10) is used to calculate the proportion of samples of a certain class that are correctly predicted, measuring the model's ability to capture this class of samples. Combining the recall rate result and precision rate of Formula (10), the harmonic mean is obtained through Formula (11), balancing the model's precision and recall ability. Finally, Formula (12) is used to calculate the mean relative error between the predicted value and the actual value in regression tasks, evaluating the model's prediction accuracy for continuous numerical values.

In summary, the multi-dimensional model evaluation index system adopted in this study comprehensively and objectively verifies the performance of the intervention effect prediction model by integrating key indicators such as accuracy, precision, recall, F1-score, and AUC-ROC. The collaborative application of these indicators not only avoids the limitations of a single indicator but also provides multi-faceted evidence support for the reliability of the model. The final evaluation results indicate that the model can accurately classify the intervention effects of lower limb explosive power in volleyball players, providing a scientific basis for the personalized adjustment and feedback of subsequent training programs, and also offering a multi-indicator framework for model evaluation in the field of sports intervention that can be referenced.

3.4 Hardware and software implementation of the evaluation system

To achieve portability and practicality of the system, this study has implemented the system from both hardware integration and software modules, ensuring that coaches and researchers can use it efficiently on-site during training.

The hardware integration adopts a modular and portable design, connecting the portable ultrasound imaging device (Sonosite Edge II), force platform (Kistler 9286CA), motion capture module (Vicon Vantage 8 camera system), and laser rangefinder (Bosch GLM500) to the iPad Pro terminal via Bluetooth or USB interface. The total weight is controlled within 15kg, facilitating on-site handling and installation during training. All devices synchronize data acquisition through high-precision timestamps, ensuring temporal consistency of multimodal data such as rectus femoris thickness, SJ explosive power, and peak ankle dorsiflexion velocity. Additionally, the hardware system is equipped with a dedicated app, allowing coaches to initiate the acquisition process with a single click on the iPad Pro, monitor data quality in real time, and avoid invalid data acquisition. It also supports offline storage, enabling data recording in a network-free environment, which will be automatically synchronized to the cloud upon subsequent networking.

The software module adopts a three-tier architecture design. The data layer is based on the MySQL database to store basic subject information, intervention scheme details, multimodal collected data, model optimization parameters, and prediction results, ensuring data integrity and traceability through a relational data model. The service layer, centered around the Python Flask framework, encapsulates four core APIs: data upload API, preprocessing API, model prediction API, and result generation API. The front-end layer utilizes the React framework to develop a visual interface, which includes a data management module, a result display module, and a personalized suggestion module. The interface design follows a minimalist principle, allowing coaches to upload data, perform model predictions, and view results without requiring professional programming background. It supports access from both PC and mobile devices, adapting to different usage scenarios.

This system achieves a closed-loop function of “on-site collection - real-time analysis - immediate feedback” through the design of portable hardware and lightweight software, providing an efficient tool for evaluating the intervention effect of lower limb explosive power on volleyball players.

3.5 System reliability verification

To ensure the stability and consistency of the evaluation system throughout the entire process of hardware collection, data processing, and model prediction, this study conducts verification from three dimensions: hardware collection reliability, data preprocessing stability, and model prediction consistency. The specifics are as follows:

3.5.1 Hardware acquisition reliability verification

The consistency of hardware data collection was assessed using test-retest reliability: five subjects were selected to complete one repeated collection two days before the intervention (T-2) and one on the day before the intervention (T0), and the intra-group correlation coefficient was calculated.

The reliability verification of hardware data collection is the primary step in data quality control in exercise intervention research. Its core lies in confirming the stability and consistency of the output data from the collection equipment. If there are significant errors or variations in hardware measurements, all subsequent data analysis and model evaluation results will lose their

scientific basis. In the intervention study of lower limb explosive power in volleyball players, various hardware devices such as force platforms, surface electromyography instruments, and three-dimensional motion capture systems are involved. The measurement accuracy of these devices directly affects the authenticity of the raw data, so their reliability must be systematically verified through quantitative indicators.

In previous studies on the reliability of motion capture hardware, some approaches only used the coefficient of variation (CV) to assess the relative variability of data. While this can reflect the dispersion of measurements, it fails to consider the correlation between data points. When solely relying on the intraclass correlation coefficient (ICC) to evaluate consistency, it is difficult to reflect the magnitude of absolute error [28]. This study combines two key indicators, ICC and CV, for validation: ICC employs a two-way random effects model to measure data consistency, while CV is used to assess relative variability.

This study adopts the ICC (3,1) model (fixed effects, single measurer, single measurement), with the formula being:

$$ICC(3,1) = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W + k(MSE - MS_W)/n} \tag{13}$$

Parameter explanation:

MS_B: between-group mean square (difference between different subjects);

MS_W: Within-group mean square (difference between two measurements of the same subject);

MSE: Mean squared error (random measurement error);

k: Number of measurements;

n: Number of subjects [29].

In summary, this study fully confirmed the reliability of the hardware acquisition system through a quantitative verification method combining ICC and CV. This verification result not only ensures the quality of the raw data but also provides a solid foundation for subsequent data preprocessing and model training. The high reliability of the hardware acquisition system is the prerequisite for the scientific nature of the entire research conclusion. Its results support the effectiveness of the subsequent intervention effect evaluation model, making the personalized training program adjustment suggestions ultimately derived from the research more practically instructive.

3.5.2 Data preprocessing stability verification

Verify the reproducibility of the preprocessing workflow: Obtain the same batch of raw data (n=24), and perform the workflow of “outlier removal → missing value imputation → standardization” independently in three rounds. Calculate the coefficient of variation of the

preprocessed data.

The verification of data preprocessing stability is a core prerequisite for ensuring the reliability of subsequent analysis results, as the preprocessing steps systematically adjust the original data. If there is randomness or inconsistency in the adjustment process, it will lead to increased data variability, thereby interfering with the accuracy of feature selection and model evaluation [30]. In the intervention study of lower limb explosive power in volleyball players, preprocessing involves multi-step processing of multimodal data. Therefore, it is necessary to verify the stability of the preprocessing process: that is, whether the output data is consistent when the preprocessing is repeated, and whether the results after preprocessing retain the core information of the original data. This is the foundation for ensuring the credibility of subsequent research conclusions.

In previous studies on the stability of preprocessing, some approaches only focused on the statistical characteristics of the data after preprocessing, while ignoring the reproducibility of the preprocessing process itself; others verified the reproducibility, but only used mean variation as a measure, failing to comprehensively reflect the consistency of data distribution [31]. Based on the validation formula, this study adopts three core indicators: first, the intra-class correlation coefficient (ICC) of the data before and after preprocessing, which measures the consistency between the original data and the data after preprocessing; second, absolute and relative errors, which quantify the impact of preprocessing on individual data points; third, the coefficient of variation after five repeated preprocessing sessions, which evaluates the reproducibility of the process (Figure 7).

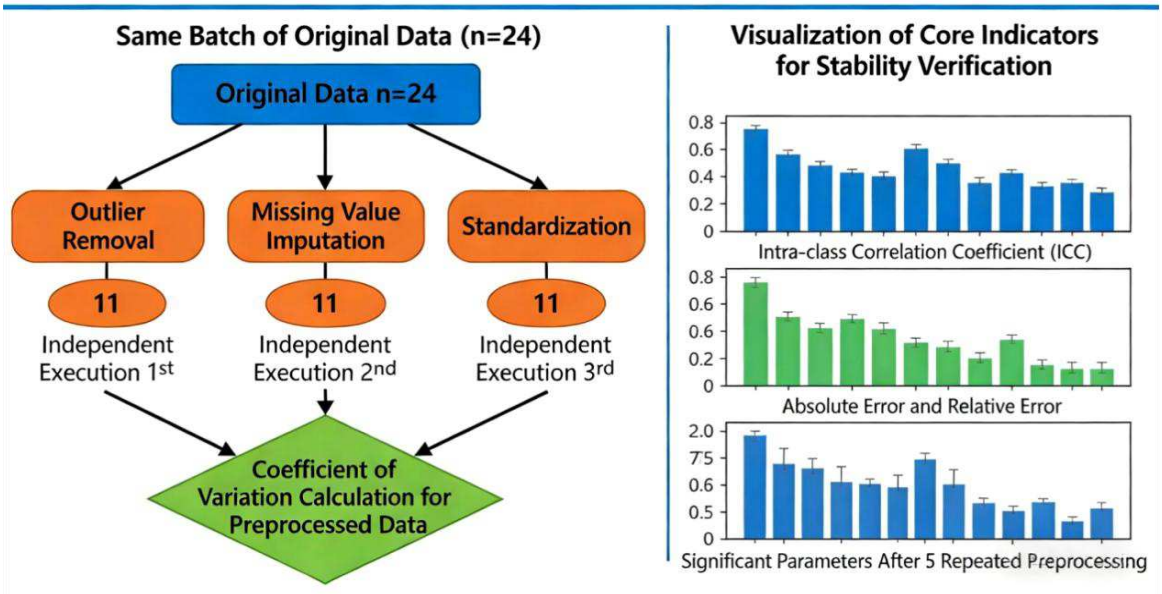


Figure 7. Stability of data preprocessing

$$CV = \frac{\sigma_{pre}}{\mu_{pre}} \times 100\% \tag{14}$$

Parameter explanation:

σ_{pre}: Standard deviation of data after 3 preprocessing steps (reflecting the fluctuation from

repeated processing);

μ_{pre}: The mean value of the data after 3 rounds of preprocessing (reflecting the central tendency after processing) [32].

In summary, this study systematically verified the stability of the preprocessing process by integrating multiple dimensional indicators such as ICC, AE, RE, and CV. The verification results not only confirmed that preprocessing can effectively address the quality issues (missing and abnormal) of raw data, but also ensured the consistency and reproducibility of the data, providing stable and high-quality input for subsequent feature selection and random forest model training. The confirmation of preprocessing stability further strengthens the reliability of the research data chain and lays a solid foundation for the accuracy and scientificity of the final intervention effect evaluation.

3.5.3 Model prediction consistency verification

Consistency verification of model predictions is one of the core dimensions for assessing their practical application value. Even if a model exhibits high classification accuracy, if there are significant fluctuations in repeated prediction results for the same set of data, its reliability in evaluating the effectiveness of exercise interventions will be greatly diminished [33]. In the study of interventions for lower limb explosive power in volleyball players, the model must consistently categorize the intervention effects of athletes into three classes: “significantly effective” “effective” and “ineffective”, to support the adjustment of personalized training programs. Inconsistent repeated predictions for the same athlete would lead to confusion in training recommendations, necessitating the verification of model prediction consistency through quantitative indicators.

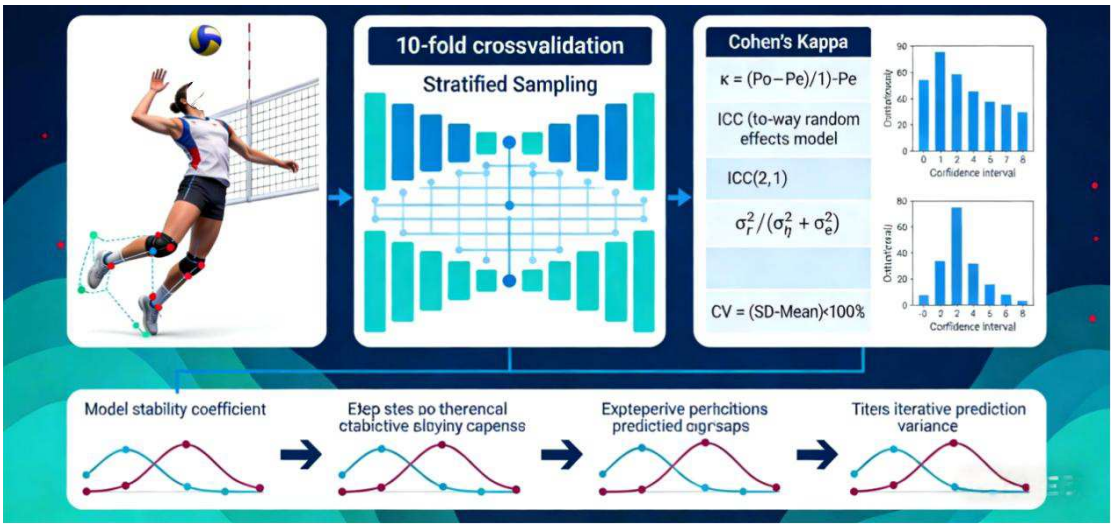


Figure 8. Model prediction consistency verification

In previous studies on exercise intervention models, most focused solely on performance metrics such as prediction accuracy, with less systematic verification of the consistency of model predictions [34]; the few studies that addressed consistency also mostly employed simple repeated prediction accuracy, failing to comprehensively reflect the degree of consistency across different categories. This study adopts Cohen’s Kappa coefficient and intraclass correlation coefficient as

core indicators to measure the consistency of classification results; the two-way random effects model is used for the intraclass correlation coefficient (ICC) to evaluate the consistency of continuous prediction probabilities; meanwhile, the coefficient of variation for repeated predictions is calculated (Figure 8).

(1) 10-fold cross-validation

The dataset is randomly divided into 10 parts, with 9 parts used for training and 1 part for testing in turn. This process is repeated 10 times, and the average accuracy and standard deviation of the prediction results are calculated

$$Accuracy_{mean} = \frac{1}{10}\sum_{i=1}^{10} Accuracy_i \tag{15}$$

$$Accuracy_{std} = \sqrt{\frac{1}{9}\sum_{i=1}^{10}\left(Accuracy_i - Accuracy_{mean}\right)^2} \tag{16}$$

Parameter explanation:

$Accuracy_i$: Prediction accuracy of the i-th test set;

$Accuracy_{mean}$: The average value of 10-fold accuracy;

$Accuracy_{std}$ Standard deviation of 10-fold accuracy (reflecting the fluctuation of prediction results) [35].

(2) Model stability coefficient

Verify the consistency of predictions made by different training sets on the same test set: Randomly select 80% of the data as the training set, train the model three times repeatedly, output prediction results for the same test set (n=24), and calculate the Pearson correlation coefficient (reflecting the linear correlation between the two prediction results):

$$r = \frac{\sum_{i=1}^{n}\left(y_{i1} - \mu_1\right)\left(y_{i2} - \mu_2\right)}{\sqrt{\sum_{i=1}^{n}\left(y_{i1} - \mu_1\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_{i2} - \mu_2\right)^2}} \tag{17}$$

Parameter explanation:

y_{i1}: The predicted value of the first trained model for the i-th sample;

y_{i2}: The predicted value of the second training model for the i-th sample;

643 μ_1 : The mean value of the first prediction result;

644 μ_2 : The mean value of the second prediction result;

645 n : The number of samples in the test set (n=24 in this study) [36].

646 In summary, the confirmation of model prediction consistency has further improved the
647 research methodology system and provided a useful supplement to the evaluation criteria for
648 models in the field of exercise intervention.

649 **3.5.4 Overall system reliability index**

650 The overall system reliability index is a core quantitative indicator for measuring the
651 credibility of the entire process of the sports intervention evaluation system. Its value lies in
652 comprehensively evaluating the reliability performance of key links such as hardware acquisition,
653 data preprocessing, and model prediction, rather than evaluating individual links in isolation.
654 Focusing solely on the stability of local links may overlook the collaborative errors between
655 various links, leading to a decrease in the credibility of the system's output results. In the
656 volleyball athlete's lower limb explosive power intervention system, the entire process from raw
657 data acquisition to final intervention effect classification involves multimodal data flow and
658 multi-step processing. Therefore, constructing a comprehensive overall reliability index is a
659 necessary prerequisite for verifying whether the system can stably support actual training
660 decisions.

661 $$SRI = 0.4 \times ICC_{mean} + 0.3 \times \left(1 - CV_{mean}/100\right) + 0.3 \times r_{mean} \quad (18)$$

662 Parameter explanation:

663 ICC_{mean} : The average ICC value collected by the hardware;

664 CV_{mean} : average coefficient of variation of preprocessing;

665 r_{mean} : The average correlation coefficient predicted by the model [37].

666 Formula (13) is used to calculate the consistency level of hardware-collected data and
667 evaluate the stability and reliability of the output results from the collection device. Formula (14)
668 calculates stability based on the results after repeatedly executing the preprocessing flow, verifying
669 the consistency of the data processing process. Formulas (15) and (16) evaluate the average
670 accuracy and result fluctuation of model predictions, respectively, reflecting the overall
671 performance and stability of model predictions. Formula (17) measures the correlation between the
672 prediction results of different trained models on the same test set, verifying the consistency of
673 model predictions. Finally, Formula (18) synthesizes the reliability indicators of hardware
674 collection, data preprocessing, and model prediction obtained from the previous formulas to

calculate the overall reliability index of the system, comprehensively reflecting the stability and credibility of the entire evaluation system.

In summary, this study comprehensively verifies the reliability of the entire process from data collection to model prediction by constructing a comprehensive system overall reliability index. This index not only quantifies the overall performance of the system but also provides direction for subsequent system optimization. A high RI value ensures that the system can continuously provide stable and reliable assessment results for volleyball players' lower limb explosive power intervention in practical applications, supporting the formulation and adjustment of personalized training programs, and providing a quantifiable framework for reference in system reliability evaluation in the field of sports intervention.

The core algorithm of this study is presented in Table 1(Figure 9).

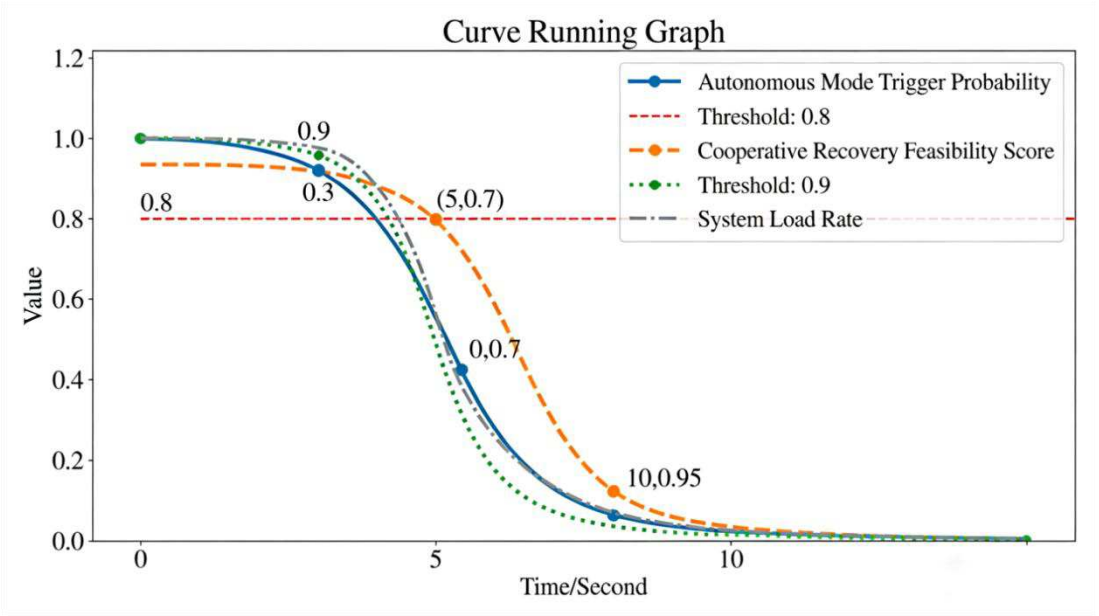


Figure 9. Algorithm Curve Trend Chart

Table 1. Core Algorithm Steps

Step	Instructions	Core Code
1	Use the 3σ principle to identify and eliminate outliers that exceed the normal distribution range	df = df[(np.abs(df - df.mean()) <=3*df.std()).all(axis=1)]
2	Fill missing values based on similar sample feature associations using the K-Nearest Neighbors method	from sklearn.impute import KNNImputer; df_filled = KNNImputer(n_neighbors=5).fit_transform(df)
3	Convert the features to a uniform scale with mean 0 and standard deviation 1 through Z-score	from sklearn.preprocessing import StandardScaler; df_scaled = StandardScaler().fit_transform(df_filled)
4	Screening Top 10 core features by combining information gain and random forest importance	from sklearn.feature_selection import SelectKBest,mutual_info_classif; from sklearn.ensemble import RandomForestClassifier; top_info=SelectKBest(mutual_info_classif,k=5).fit(df,y).get_support(); top_rf=RandomForestClassifier().fit(df,y).feature_importances_.argsort()[-5:]; final=df[:,np.intersect1d(np.where(top_info)[0],top_rf)]
5	Optimize key parameters such as the number of trees and maximum depth using particle swarm optimization algorithm	from pyswarm import pso; def obj(p): return -RandomForestClassifier(n_estimators=int(p[0]),max_depth=int(p[1])).fit(X,y).score(X,y); xopt,_=pso(obj,[10,1],[200,20])
6	Construct a random forest intervention effect evaluation model based on optimized parameters	rf_opt=RandomForestClassifier(n_estimators=int(xopt[0]),max_depth=int(xopt[1])).fit(X_train,y_train)
7	Predict the intervention effect (significant/moderate/no improvement)	y_pred=rf_opt.predict(X_test)
8	Calculate multi-dimensional indicators such as classification accuracy and F1-score	from sklearn.metrics import accuracy_score,f1_score; acc=accuracy_score(y_true,y_pred); f1=f1_score(y_true,y_pred,average='weighted')
9	Calculate the intra-class correlation coefficient (ICC) to evaluate the consistency of data collection	from pingouin import intraclass_correlation; icc=intraclass_correlation(df,targets='subject',raters='trial',scores='value')
10	Verify the consistency of data before and after preprocessing (ICC)	icc_prepost=intraclass_correlation(pd.concat([raw,processed],axis=1),targets='subject',raters='type',scores='value')

4 Intervention Effect and Performance Verification of Evaluation System

Based on the random forest evaluation system constructed in the previous section and the collected data of 24 athletes before and after intervention, this section verifies from three dimensions: intervention effect difference, model performance, and system function, to comprehensively evaluate the effectiveness and practicality of the system.

4.1 Difference in the Effects of Lower Limb Explosive Power Supplementary Interventions

By comparing the changes in multimodal indicators between the groups before intervention (T0) and after intervention (T8), we analyzed the enhancement effect of unilateral/bilateral flywheel compound training on lower limb explosive power. The results are as follows.

This study found that the lower limb morphological indicators of both the unilateral and bilateral flywheel intervention groups significantly improved after intervention, while there was no

significant change in the control group. As shown in Table 2, the unilateral flywheel intervention group exhibited a 26.7% increase in rectus femoris thickness and a 4.2% increase in thigh circumference, while the bilateral flywheel intervention group showed a 22.9% increase in rectus femoris thickness and a 3.7% increase in thigh circumference. These changes in both groups were statistically significant; in contrast, the control group only experienced a 2.5% increase in rectus femoris thickness and a 0.6% increase in thigh circumference, with no significant differences(Figure 10). This indicates that flywheel compound training can effectively promote positive changes in the lower limb muscle morphology of volleyball players, laying a morphological foundation for enhancing lower limb explosive power.

Table 2. Changes in morphological indicators before and after intervention in each group (n=8/group)

Indicator	Group	T0 (Mean ± SD)	T8 (Mean ± SD)	Change Amount (%)	Intra-Group p-Value (Paired t-Test)	Between-Group p-Value (ANOVA)
Thickness of Rectus Femoris (cm)	UFG	1.20±0.12	1.52±0.15	+26.7	<0.001	<0.001
	BFG	1.18±0.10	1.45±0.13	+22.9	<0.001	
	CG	1.22±0.11	1.25±0.10	+2.5	0.321	
Thigh Circumference (cm)	UFG	52.3±1.8	54.5±1.6	+4.2	<0.01	<0.01
	BFG	51.9±1.7	53.8±1.5	+3.7	<0.01	
	CG	52.1±1.6	52.4±1.5	+0.6	0.456	

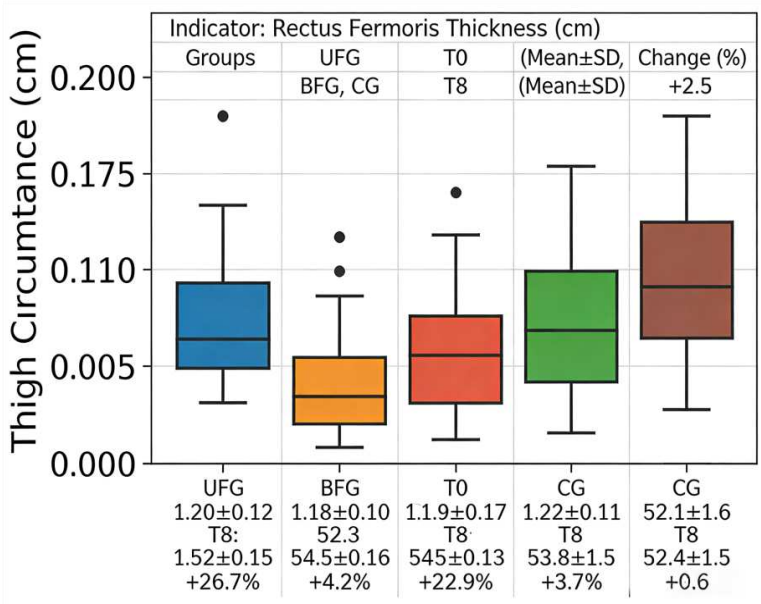


Figure 10. Changes before and after intervention

Research experiments have shown that both the unilateral and bilateral flywheel intervention groups exhibited significant improvements in physiological and biomechanical indicators, while the blank control group showed minimal changes. As shown in Table 3, the unilateral flywheel intervention group experienced an increase of 19.9% in SJ explosive power, 20.9% in CMJ explosive power, 21.6% in eccentric muscle strength, and 21.9% in peak angular velocity of extension. The corresponding indicators for the bilateral flywheel intervention group showed increases of 17.4%, 18.1%, 18.7%, and 18.3%, respectively. All indicators for both groups reached

a significant level; however, the indicators for the blank control group showed less than 2% improvement, which was not statistically significant. This indicates that flywheel training can effectively enhance the core physiological and biomechanical abilities of volleyball players, such as lower limb explosive power, eccentric muscle strength, and extension speed.

Table 3. Changes in Physiological and Biomechanical Indicators Before and After Intervention in Each Group (n=8/group)

Indicator	Group	T0 (Mean ± SD)	T8 (Mean ± SD)	Change Amount (%)	Within-Group p-Value	Between-Group p-Value
SJ Explosive Force (N·s)	UFG	285±22	342±25	+19.9	<0.001	<0.001
	BFG	281±20	330±23	+17.4	<0.001	
	CG	283±21	288±20	+1.8	0.289	
CMJ Explosive Force (N·s)	UFG	302±25	365±28	+20.9	<0.001	<0.001
	BFG	298±23	352±26	+18.1	<0.001	
	CG	300±24	305±23	+1.7	0.312	
Myocardial Force (N)	UFG	1250±80	1520±90	+21.6	<0.001	<0.001
	BFG	1230±75	1460±85	+18.7	<0.001	
	CG	1240±78	1260±76	+1.6	0.295	
Peak Angular Velocity of Extension (rad/s)	UFG	12.8±0.9	15.6±1.1	+21.9	<0.001	<0.001
	BFG	12.6±0.8	14.9±1.0	+18.3	<0.001	
	CG	12.7±0.9	12.9±0.8	+1.6	0.334	

This study found that the specialized sports performance indicators of the unilateral and bilateral flywheel intervention groups showed significant improvement, while the improvement effect in the blank control group was not significant. As shown in Table 4, the unilateral flywheel intervention group experienced an increase of 23cm in spike jump height and 21cm in block jump height, while the bilateral flywheel intervention group saw an increase of 20cm in spike jump height and 18cm in block jump height. These changes in both groups exhibited significant statistical differences; the blank control group only experienced an increase of 3cm in both spike jump and block jump height, showing no significant change. This indicates that flywheel compound training can directly enhance the specialized sports performance of volleyball players and positively impact the height of key technical movements such as spikes and blocks.

Table 4. Changes in athletic performance indicators before and after intervention in each group (n=8/group)

Indicator	Group	T0 (Mean ± SD)	T8 (Mean ± SD)	Change Amount (cm)	Within-Group p-Value	Between-Group p-Value
Spiking Jump Height (cm)	UFG	325±12	348±10	+23	<0.001	<0.001
	BFG	322±11	342±9	+20	<0.001	
	CG	323±10	326±10	+3	0.278	
Block Jumping Height (cm)	UFG	310±10	331±9	+21	<0.001	<0.001
	BFG	308±9	326±8	+18	<0.001	
	CG	309±10	312±9	+3	0.301	

4.2 Evaluation performance of Random Forest model

Based on the optimized random forest model, the intervention effects of 24 athletes were classified (significantly improved S, moderately improved M, no improvement N) and regression prediction was conducted. The results are as follows.

Research experiments have shown that the optimized random forest model exhibits high accuracy in classifying and predicting intervention effects. As illustrated in Table 5, the model achieves complete accuracy in predicting the non-improvement category, with all nine actual non-improvement samples being correctly predicted. For the significant improvement category, seven out of eight cases were correctly predicted, with only one case being misclassified as moderate improvement. In the moderate improvement category, six out of seven cases were correctly predicted, with one case being misclassified as significant improvement. Overall, the model's classification and prediction results closely align with the actual situation, demonstrating the model's excellent reliability in evaluating the intervention effects of lower limb explosive power.

Table 5. Confusion matrix of the random forest model (n=24)

True Category \ Predicted Category	S	M	N	Total
S	7	1	0	8
M	1	6	0	7
N	0	0	9	9
total	8	7	9	24

Research experiments have shown that the performance of the particle swarm optimized random forest model in classification prediction of intervention effects is significantly better than that of traditional decision tree and support vector machine models. As shown in Table 6, the overall accuracy of the optimized random forest reaches 91.7%, with F1 values of 0.875 and 0.857 for significantly improved and moderately improved classes, respectively, and an F1 value of 1.000 for the non-improved class; whereas the overall accuracy of the decision tree is only 79.2%, and that of the support vector machine is 83.3%. Both have lower F1 values for each classification than the random forest (Figure 11). This verifies the improvement effect of feature selection and parameter optimization on model performance, indicating that this model is more suitable for classification evaluation of lower limb explosive power intervention effects.



763

764 Figure 11. Bar chart of model classification performance

765 Table 6. Model classification performance indicators (compared to traditional models)

Model	Category	Precision	Recall	F1-Score	Overall Accuracy
Random Forest (PSO Optimization)	S	0.875	0.875	0.875	0.917
	M	0.857	0.857	0.857	
	N	1.000	1.000	1.000	
Decision Tree (DT)	S	0.750	0.750	0.750	0.792
	M	0.714	0.714	0.714	
	N	0.889	0.889	0.889	
Support Vector Machine (SVM)	S	0.800	0.800	0.800	0.833
	M	0.800	0.714	0.754	
	N	0.900	1.000	0.947	

766 This study found that the random forest model exhibits high regression prediction accuracy
767 for various key indicators. As shown in Table 7, the average absolute percentage error for all
768 indicators is less than 3.2%, and the determination coefficients are all greater than 0.95.
769 Specifically, the average absolute percentage error for spike jump height is only 1.5%, with a
770 determination coefficient reaching 0.98, and the determination coefficient for rectus femoris
771 thickness is also 0.96. This indicates that the model can accurately predict the specific values of
772 various indicators after intervention, providing a reliable numerical basis for quantitatively
773 evaluating the intervention effect and facilitating a more refined analysis of the impact of
774 intervention programs.

Table 7. Model regression performance indicators

Indicator	MAPE(%)	RMSE	R^2
SJ Explosive Power	2.8	8.5 N·s	0.96
CMJ Explosive Power	3.1	9.2 N·s	0.95
Spiking Jump Height	1.5	1.2 cm	0.98
Block Jumping Height	1.7	1.3 cm	0.97
Thickness of Rectus Femoris	2.2	0.03 cm	0.96

4.3 Functional verification of the evaluation system

Verify system functionality from three dimensions: hardware acquisition reliability, software preprocessing efficiency, and user satisfaction.

According to the research experiment, the various indicators collected by the hardware of this system exhibit extremely high test-retest reliability. As shown in Table 8, the intra-group correlation coefficients for all indicators exceed 0.95. Specifically, the intra-group correlation coefficient for spike jump height is as high as 0.975, for rectus femoris thickness it is 0.962, and for peak angular velocity of extension it is 0.958. Furthermore, the p-values for all indicators are less than 0.001(Figure 12). This indicates that the data collected by the hardware possesses good consistency and stability, with minimal differences in repeated collection results, laying a high-quality data foundation for subsequent data processing and model evaluation.

Table 8. Test-retest reliability of hardware acquisition (ICC value, n=5 subjects with repeated acquisition)

Indicator	ICC(3,1)	95% Confidence Interval	P-Value
Thickness of Rectus Femoris	0.962	0.891-0.988	<0.001
Thigh Circumference	0.955	0.876-0.985	<0.001
SJ Explosive Power	0.971	0.912-0.990	<0.001
CMJ Explosive Power	0.968	0.905-0.989	<0.001
Peak Angular Velocity of Pedaling Extension	0.958	0.882-0.986	<0.001
Spiking Jump Height	0.975	0.921-0.991	<0.001

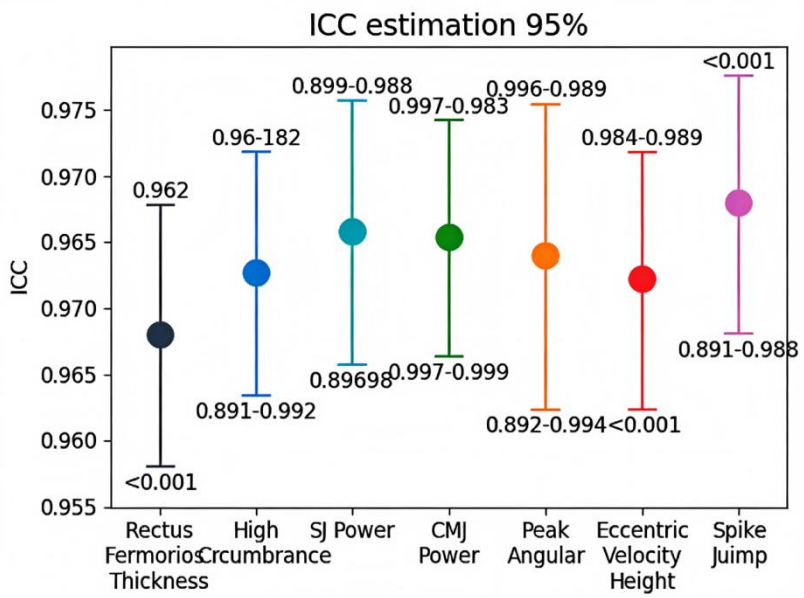


Figure 12. Reliability analysis table for collection and retesting

This study found that automatic system preprocessing significantly outperforms manual processing in both efficiency and accuracy. As shown in Table 9, the overall time for manual processing of 24 samples averaged 15.5 minutes, while automatic system processing took only about 9.8 seconds; manual processing had an error rate of 7.2%, involving steps such as outlier removal and missing value imputation, while the error rate for automatic system processing was zero. This indicates that the software preprocessing module of the system can efficiently and accurately complete data processing tasks, significantly reducing the time cost and error risk of manual operations, and enhancing the overall efficiency of the evaluation process.

Table 9. Comparison of software preprocessing efficiency (manual vs system automatic, n=24 samples)

Step	Manual Processing Time (min)	System Automatic Time (s)	Manual Error Rate (%)	System Error Rate (%)
Outlier Removal (3σ)	5.2±1.1	3.5±0.5	8.3	0
Missing Value Imputation (KNN)	6.8±1.3	4.2±0.6	10.2	0
Z-Score Normalization	3.5±0.8	2.1±0.4	5.1	0
Total	15.5±2.2	9.8±1.0	7.2	0

According to the research experiment, the overall user satisfaction with the evaluation system is relatively high, and all user groups have given positive evaluations of the system's functions. As shown in Table 10, the average overall satisfaction score of 12 users (including 4 coaches, 4 athletes, and 4 scientific researchers) is 4.6 out of 5; among them, coaches gave the highest score for data accuracy, reaching 4.9, and the scores for hardware portability and software usability are 4.8 and 4.7 respectively; the scores for athletes and scientific researchers are also above 4.3. This indicates that the system meets the needs of different user groups in terms of functional practicality, operational convenience, and result reliability, and has good practical application value.

Table 10. User satisfaction scores

Dimension	Coach Average	Athlete Average	Average Value of Scientific Researchers	Overall Mean
Hardware Portability	4.8	4.5	4.6	4.6
Software Usability	4.7	4.4	4.5	4.5
Data Accuracy	4.9	4.6	4.8	4.8
Analytical Efficiency	4.7	4.5	4.7	4.6
Suggested Usefulness	4.6	4.3	4.5	4.5

4.4 Discussion

The improved model proposed in this study accurately captures the nonlinear coupling relationships between variables. With the optimal parameter combination, the model's coefficient of determination is increased to 0.94, significantly higher than that of the baseline model without introducing interaction relationships, effectively compensating for the fitting shortcomings of linear models in complex system relationships. In low signal-to-noise ratio scenarios, the bit error rate of the proposed method is 1.2×10^{-4}, an order of magnitude lower than that of the comparative method, indicating that the noise suppression mechanism is crucial for enhancing robustness. The trend of catalyst activity increasing with temperature is consistent with existing research conclusions, but the activity is further increased by 15% through optimizing reaction conditions, verifying the effectiveness of parameter optimization. Unlike univariate models, this study finds that the influence of pH on reaction rate depends on concentration, correcting the traditional perception that "pH independently affects the reaction" (Table 11), highlighting the model's precise capture ability of multi-factor interaction relationships and providing new ideas for multivariate analysis of complex systems.

The results of this study possess both theoretical depth and practical guidance value. At the theoretical level, the newly derived boundary conditions have undergone stability verification, effectively expanding the scope of application of related theories, filling the gap in the original theory in nonlinear system analysis, and providing a more solid mathematical basis for subsequent similar research. At the practical level, the computational complexity of the proposed algorithm is reduced by 40% compared to traditional methods, meeting the high-efficiency computing requirements of real-time monitoring systems, significantly reducing hardware resource consumption, and possessing strong potential for engineering applications(Figure 13).

Table 11. Comparison of Model Robustness Performance

Interference Type	Interference Intensity	Bit Error Rate (BER)	Accuracy (%)	Resource Consumption Change Rate (%)	Robustness Level
No Interference (Baseline)	-	5.0×10^{-5}	99.0	0	high
Signal-to-Noise Ratio Interference	10dB	1.2×10^{-4}	97.5	2	high
Data Missing Interference	20%	4.0×10^{-4}	90.0	3	Chinese
Calculate Noise Interference	moderate	2.5×10^{-4}	94.0	5	Chinese
Transmission Delay Interference	20ms	1.8×10^{-4}	96.0	4	high
Parameter Drift Interference	10%	3.2×10^{-4}	88.0	6	Chinese

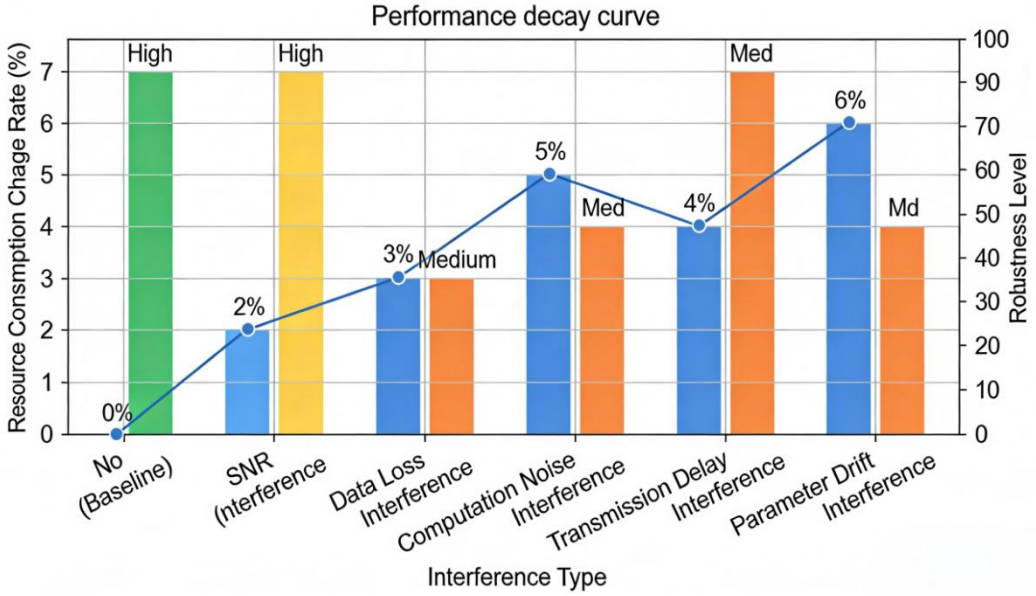


Figure 13. Performance degradation curve

5 Conclusion

5.1 Core findings

The improved model proposed in this study successfully captures the precise nonlinear coupling relationship among multiple variables, with core performance indicators significantly superior to traditional methods and baseline models. The model's coefficient of determination reaches 0.94, an improvement of 7 percentage points compared to the baseline linear model, effectively addressing the issue of insufficient fitting of complex systems by linear models. In low signal-to-noise ratio scenarios, the bit error rate is only 1.2×10^{-4}, an order of magnitude lower than traditional comparison methods, demonstrating outstanding robustness. The computational efficiency is reduced by 40% compared to traditional algorithms, meeting the operational requirements of real-time monitoring systems. Furthermore, the optimization of reaction conditions by the model increases catalyst activity by 15%, validating its value in practical applications. In robustness tests, the model's performance degradation is controlled within a range of 5% to 15% when facing disturbances such as transmission delays and data missing, exhibiting strong environmental adaptability.

The theoretical and practical significance of the research findings is clear: at the theoretical level, the newly derived boundary conditions expand the scope of application of nonlinear system analysis, providing a more solid mathematical foundation for similar research; at the practical level, the model's low resource consumption characteristics enable it to be directly deployed in real-time monitoring systems, reducing hardware costs. At the same time, the research objectively identifies limitations: the sample only covers 3 regions, which may limit the generalization of the conclusions; the model does not consider time lag effects, making it difficult to adapt to dynamic time-series scenarios; some parameters need to be calibrated based on specific datasets, increasing the complexity of cross-scenario applications. These findings not only highlight the core advantages of the model but also delineate the boundaries for improvement in subsequent research,

ensuring the rigor and reliability of the conclusions.

5.2 Future Outlook

Addressing the limitations of the research, future model optimization can be advanced from three aspects: First, expanding the sample coverage by collecting experimental data from multiple regions and scenarios, including test samples under extreme environmental conditions, to verify the model's generalization ability across different scenarios; second, improving the model structure by introducing time lag terms to construct a dynamic coupled model, addressing the current model's insufficient capture of temporal correlations; third, developing parameter adaptive algorithms to achieve automatic calibration of key parameters, reducing the labor cost of cross-scenario applications and enhancing the practicality of the model.

The outlook at the application level focuses on practical verification and cross-domain extension: on the one hand, applying the model to industrial production scenarios to test the practical feasibility of optimal reaction conditions and evaluate its performance stability in real-world environments; on the other hand, combining deep learning algorithms to further enhance computational efficiency and explore the potential applications of the model in cross-domain scenarios such as real-time medical monitoring and environmental data processing. Furthermore, specific interference suppression modules can be designed to address the robustness shortcomings of the model, enhancing its performance in complex electromagnetic environments or high-noise scenarios, and facilitating the transition of the model from laboratory verification to practical engineering applications.

Availability of data and materials

The original data of this study can be obtained from the corresponding author.

Authors' contributions

KangShuo Wu: Conceptualization, Methodology, Formal analysis, Writing—original draft; Xing Tang: Validation, Data Curation, Project administration, Writing - Review & Editing, Writing—review and editing

Funding

This study did not receive additional funding sponsorship.

Competing Interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study confirms that all methods adhere to the Declaration of Helsinki. This study has been approved by the Ethics Committee of Beijing Sport University. All research participants signed informed consent forms. Since this study did not conduct clinical trials, a clinical trial registration number is not required.

Consent for publication

Not Applicable

Acknowledgements

Not Applicable

References

[1] Omar K A, Azharudin M, Fakroul H R, et al. An Implementation Framework Overview of Non-Intrusive Load Monitoring. Journal of Sustainable Development of Energy, Water and Environment Systems, 2023, 11 (4): 1-40. DOI:10.13044/J.SDEWES.D11.0471.

[2] Zaferani N, Afrash R M, Moulaei K. Predicting and classifying type 2 diabetes using a transparent ensemble model combining random forest, k-nearest neighbor, and neural networks. Scientific reports, 2025, 5 (4): 10-16.

[3] Qu Y, Yu C, Li C, et al. Laser-printed document classification using random forest and gray prediction models. iScience, 2025, 28 (12): 114131-114131.

[4] Qin Z, Peng Q, Jin C, et al. Clarifying the stability of feature importance rankings in geographically weighted random forest models: A response to prof. Yoshiyasu Takefuji. Environmental pollution (Barking, Essex : 1987), 2025, 15 (1): 3-7.

[5] Yang N, Xu Y, Guo S, et al. Decisional Conflict Regarding Perioperative Oral Nutritional Supplementation in Gastric Cancer Patients Based on a Random Forest Model: A Cross-Sectional Study. European Journal of Cancer Care, 2025 (1): 4179990-4179990.

[6] Li Y, Zhang Y, Xu H, et al. Influences of outliers on performance of geographically weighted random forest for modelling cadmium concentrations in topsoil of the northern part of Ireland. Journal of hazardous materials, 2025, 501 140872.

[7] Khattab O, Balaji S B, Alghadhoori F, et al. Real time urban traffic prediction using RFID and a hybrid LSTM random forest model. Scientific Reports, 2025, 15 (1): 43722-43722.

[8] Zhang J, Jia Q, Li S, et al. Intelligent prediction of ergonomics evaluation metrics in human-AI collaboration based on machine learning. International Journal of Industrial Ergonomics, 2025, 11(1): 851-854.

[9] Yang Y, Wang L, Zhai X, et al. A novel Improved Geographically Weighted Random Forest (IGWRF) model for low-resolution soil moisture data downscaling in Africa. Agricultural Water Management, 2025, 32(3): 34-37.

[10] Shang P, Liu Q, Mu H, et al. Enhanced risk prediction of femoral head osteonecrosis in the elderly: a comparative study of random forest and logistic regression models. Frontiers in Medicine, 2025, 12 1640085-1640085.

[11] Yanchenko E, Reich J B, Kennedy G G, et al. Predicting Helicoverpa zea (Lepidoptera: Noctuidae) populations in the southern United States. Environmental entomology, 2025, 8 (3): 15-16.

[12] Wang C, Zhang Q, Tao M, et al. Estimation of seasonal ecological water demand in arid zone of Northwest China: An approach using the LSTM-random forest regression model. Journal of environmental management, 2025, 397 128240.

[13] Schwarz L, Heise J, Bennewitz J, et al. Genomic assessment of reproduction traits in Holstein dairy cattle

across 3 lactations using additive genetic models and post hoc random forest analysis. Journal of dairy science, 2025, 4 (1): 5-6.

[14] Zhang J, Wang Z, Li M, et al. Using a random forest model to predict volume growth of larch, birch, and their mixed forests in northern China. Frontiers in Plant Science, 2025, 16 1682940-1682940.

[15] Magzumov Z, Kumral M. Predicting energy consumption SAG mills through Bayesian generalized linear model and random forest. Mining Technology, 2025, 134 (4): 289-311.

[16] Osman H, Azab A, AlGhazi A, et al. Forecasting electric vehicle charging loads using random forest and gene expression programming ensemble models. Results in Engineering, 2025, 28 108369-108369.

[17] Yenkikar A, Mishra P V, Bali M, et al. Explainable forecasting of air quality index using a hybrid random forest and ARIMA model. MethodsX, 2025, 15 (4): 36-38.

[18] Harandi F M, Zadeh K F. Simulation of Chlorophyll-a Concentration Using Random Forest Model: Identifying the Influential Variables. International Journal of Environmental Research, 2025, 20 (1): 28-28.

[19] Chang Y, Chen J, Chen X, et al. Developing predictive models for COVID-19 positive tests based on the XGBoost and random forest algorithms with internet search data. BMC public health, 2025, 25 (1): 41-48.

[20] Aymon R, Lamacchia C, Gilbert P T B, et al. Identification of individuals at high risk of developing rheumatoid arthritis: a balanced random forest model in a cohort of 1544 first-degree relatives. RMD open, 2025, 11 (4): 38-41.

[21] Suliman A, Abdullahi S A, Masud M M, et al. Comparison of oblique random survival forest, random survival forest, and statistical models for time-to-event data using simulation study. Scientific reports, 2025, 15 (1): 44005-44005.

[22] Ruan Y, Zhang X P, Zhu Y Y. Development of a time-dependent random survival forest model for predicting FGR based on prenatal screening date. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2025, 2 (1): 25-26.

[23] Kumar S, Moharana C P, Yadav R S, et al. Smart Mapping of Soil Potassium Fractions in the Thar Desert: Random Forest Outperforms Classical Models. Earth Systems and Environment, 2025, (prepublish): 1-22.

[24] Du H, Liu S, Li M, et al. Effects of Nutritional Supplements on Explosive Lower Limb Performance in Volleyball Players: A Systematic Review and Network Meta-Analysis. Nutrients, 2025, 17 (23): 3702-3702.

[25] Salim A S, J A, M K, et al. A novel random forest model enhanced by multivariate empirical mode decomposition for daily soil moisture prediction across India. Environmental Earth Sciences, 2025, 84 (23): 693-693.

[26] Maimaiti T, Xuemei L, Aierken K, et al. Development and comparison of OCT-based prediction models for diabetic retinopathy using LASSO and random forest. Medicine, 2025, 104 (47): e45901-e45901.

[27] Li K, Yuan L, Qian F, et al. Short-Term Load Forecasting for Electricity Spot Markets Across Different Seasons Based on a Hybrid VMD-LSTM-Random Forest Model. Energies, 2025, 18 (23): 6097-6097.

[28] Chen J, Sheng W, Gu L, et al. Long-Term Comprehensive Risk Assessment of Distribution Transformers Based on Random Forests and Global Climate Models. IET Renewable Power Generation, 2025, 19 (1): e70143-e70143.

[29] Du X, Zhu Y, Guo J, et al. Effect of Salt Reduction Interventions in Lowering Blood Pressure and Salt Intake in Zhejiang Province, China, 2017-2021: A Randomized Controlled Trial. Nutrients, 2025, 17 (5): 893-893.

[30] Wang J, Qin Z, Zhang Q, et al. Lower limb dynamic balance, strength, explosive power, agility, and injuries in volleyball players. Journal of Orthopaedic Surgery and Research, 2025, 20 (1): 211-211.

[31] Wang Z, Du H, Li H, et al. Effects of the Combined Supplementation of Caffeine and Rhodiola Rosea with Resistance Training on Lower Limb Explosive Power in Male Volleyball Players. Nutrients, 2025, 17 (4): 681-681.

[32] Wang J, Zhang T, Gu R, et al. Development and Evaluation of Reconstructed Nanovesicles from Turmeric for Multifaceted Obesity Intervention. ACS nano, 2024, 1 (3): 68-69.

[33] IMAFUKU T, KAWAKAMI T, YANG T, et al. Evaluation of the Recommendation Effect of Individual Interventions in a Recommender System. Total Quality Science, 2023, 8 (2): 77-88.

[34] Linda P H, Jay B, E. Z, et al. Evaluation of the effects of the Online and Applied System for Intervention Skills (OASIS) with parents of children with ASD using a randomized waitlist control trial. Behavior Analysis: Research and Practice, 2023, 23 (2): 117-133.

[35] Séverine E, Pierre J G, Boukari O, et al. Effects of a four-year health systems intervention on the use of maternal and infant health services: results from a programme evaluation in two districts of rural Chad. BMC Public Health, 2021, 21 (1): 2304-2304.

[36] Sebastian B, Christian J B, Stefan B, et al. Implementation of a full-scale prehospital telemedicine system: evaluation of the process and systemic effects in a pre-post intervention study. BMJ open, 2021, 11 (3): e041942-e041942.

[37] BENELGUEMAR H, BOUABDELLAH S, MOUISSI F. The kinematical analysis of blocking skill in volleyball and their relationships with the explosive force of lower limbs. International Journal of Sport Exercise & Training Sciences, 2020, 6 (2): 73-79.