J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 2947--2965

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Using data mining techniques to optimise athlete training and recovery programmes in tertiary physical education

Yukun Lu[1,✉], Zexi Sun[1]

[1] *College of Physical Education, Xinjiang Hetian Normal College, Hetian, Xinjiang, 848000, China*

ABSTRACT

Aiming at the demand for scientific training of athletes in college sports education, this paper integrates data mining technology to propose athlete training and optimisation methods, and constructs an athlete training quality monitoring system and intelligent recovery assessment system. The traditional Apriori algorithm is improved by using multidimensional association rules, and multidimensional attribute mining is carried out on the collected data of athletes' training data to search for frequent item sets and output strong association rules, so as to achieve the monitoring of training quality and adjustment of training programmes. Using the improved fuzzy decision-making method to filter out the optimal feature subset, and integrating the improved whale algorithm and random forest to achieve intelligent recovery effect evaluation. By carrying out the practice of training and recovery optimisation, it can be seen that the total score of physical fitness test of track and field athletes increased from 18.19 to 19.8 before the experiment, and the training quality was significantly improved. Various health indicators such as heart rate, blood lactate, serum creatine kinase, etc. gained significant improvement in adopting the recovery optimisation method of athletes in this paper. The mean values of training status, coaching factors, and personal situation satisfaction evaluation dimensions were 4.35, 4.425, and 4.38, respectively, and the training and recovery plan of this experiment was well received by the subject athletes.

Keywords: apriori algorithm, multidimensional association rules, whale algorithm, random forest, athlete training

1. Introduction

The model of interaction between physical education and athletic training in higher education is a key topic in education and sport today. In the past, physical education and sports training were often regarded as two separate fields, each pursuing different goals and methods [1-3]. However, with the development of society and the evolution of educational concepts, people have gradually recognised

✉ Corresponding author.
E-mail address: m18799357033@163.com (Y. Lu).
Received 03 March 2024; Revised 12 April 2024; Accepted 15 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-167

the close connection and interaction between the two. Physical education in colleges and universities aims to cultivate students' physical quality and comprehensive literacy, while sports training focuses on improving athletes' competitive level and skills [4-6]. In this context, the interactive mode of physical education and sports training in colleges and universities has given rise to a series of innovations and reform methods, which provide broader development opportunities for students and athletes, and also bring new challenges and opportunities for higher education and the sports industry [7-8].

In the field of sports competition, the amount of data involved in each sport is very large, but as of now, the utilisation rate of these massive data is very low, and it is difficult to produce satisfactory results [9-10]. Therefore, how to improve the utilisation rate of the data and mine the potentially useful information such as athletes' training rules in these massive data has become the focus of research in the field of sports and athletics. With the vigorous development of information technology, data mining technology, as a cross-cutting discipline involving many fields, has gradually become an effective way for people to integrate and mine the useful information and rule associations in massive data [11-13]. In view of this, a study is conducted on the application of data mining technology in multi-attribute training of athletes, in order to effectively mine the potential rules and associations in the data related to athletes' training through the application of appropriate data mining technology, and filter out the valuable information therein, so as to realise the data support for athletes' training, and make the training more scientific and efficient [14-15].

The interaction mode of physical education and sports training in colleges and universities is a complex and important field, in which the curriculum and teaching methods are the core elements, and the innovation and coordination of these two aspects are crucial for cultivating students with all-round development and improving the athletes' competitive level. Literature [16] examined the practice of VR technology in the field of physical education and sports training, and concluded that the scientific and rational use of virtual reality technology has positive significance for the improvement of physical education and sports training. Literature [17] explains that mass sports activities are often neglected in the practice of physical education, and reveals the existence of such problems as not paying attention to the opinions of students, and the lack of coordination between the load of sports and the physical quality of students in the practice of physical education and sports. Literature [18] reviewed the research literature related to the development of youth physical education and individual sports training and identified the themes of this research direction, which provides a positive reference for practitioners and researchers in the field of physical education and sports training. Literature [19] attempted to optimise physical education teaching based on agile management strategies and conducted teaching experiments to demonstrate this, confirming that agile management strategies play a positive role in physical education practice. Literature [20] critically analysed the performance of multivariate training programmes applied in physical education courses, and the study helped to further develop the practical potential of multivariate training programmes, and pointed out that improving athletic proficiency and creativity is the best direction for the optimisation of multivariate programmes. Literature [21] examined the technical performance of the small-sided game (SSG) training strategy based on the literature analysis method, and the study deepened the understanding of SSG training programme among the relevant stakeholders, and provided an important reference for football coaches in the development of training programmes.

As the level of competitive sports continues to improve, the training quality of athletes has become a key competitive element. By analysing the data in the training process through data mining technology, the key factors to improve the quality of training can be identified, so as to provide coaches and athletes with scientific training suggestions. Literature [22] discusses the decision tree algorithm, which is the core logic in the Psychological Scale of Sports Performance, as a case study. The decision tree algorithm is used to evaluate the association of the factors examined in the athletes with gender and age, and points out that the decision tree can be used for predicting the future athletic data and assumptions of the sports athletes. Literature [23] examined the performance of an analytical model based on machine learning algorithms in predicting high intensity impacts suffered by youths in youth football training, and the study affirmed the good predictive performance of the ML model. Literature [24] conceived a prediction framework based on lattice clustering algorithm by combining data mining techniques and discrete Morse theory to achieve the prediction of physical fitness of sports athletes and college physical education students. Literature [25] conceived MobilePerfMiner, a data mining framework for mobile processors, and uncovered assessment methods for tennis sports training as well as characteristics of tennis technical parameters around the characteristics of the tennis sport, which helps to improve the tennis players' tennis performance. Literature [26] developed a hierarchical machine learning prediction model, which can achieve the prediction of athletes' injuries based on sports load data, and validated it using 21 football players' relevant data, which confirms the effectiveness of the proposed prediction model, and highlights the role of the model in early sports injury intervention. Literature [27] systematically reviewed the development and current status of intelligent sports training, focused on the logic and practice of intelligent data analysis methods in the field of intelligent sports training, and finally also looked forward to the trends and directions of the future development of intelligent sports training.

Combined with data mining technology, this study proposes an optimisation method for training and recovery of college sports athletes, and builds up an athlete training quality monitoring system and an intelligent recovery assessment system respectively. The traditional Apriori association rule algorithm is improved, and the multidimensional attribute mining of the collected data using the improved association rule algorithm solves the problem of in-depth analysis of the athletes' training and competition data to achieve the monitoring of the training quality and to adjust the athletes' training programme on this basis. The gait features of the athletes were extracted from the digital human motion model using the skeletal key points captured by Kinect sensor, and the optimal feature subset was screened using the improved fuzzy decision-making method, and the optimal feature subset was inputted into the intelligent recovery assessment model incorporating the random forest and improved whale algorithm, so as to realise the intelligent recovery assessment of the athletes. Twenty track and field athletes from University H of Jiangxi Province were selected as experimental subjects to carry out the practice of athlete training and recovery optimisation, and the performance tests of the athlete training quality monitoring system and the intelligent recovery assessment system constructed in this paper were carried out respectively, and the discussion and analysis around the effect of training optimisation, the effect of recovery optimisation, and the athlete's satisfaction with the training and recovery plan during the experimental period were carried out.

2. Challenges in the development of athletes for physical education in higher education

As a part of college education, the training effect of college athletes not only affects the quality of undergraduate education, but also plays an important role in the development of national competitive sports. Athletic training and recovery is the basic and necessary link needed to cultivate athletes in college sports education, and it is also the key to ensure athletes' special physical quality and special technical ability. With the rapid development of competitive sports, the requirements of athletes' training indexes are constantly improving, and colleges and universities are facing brand-new challenges in the training link of sports training and recovery.

1) Lack of effective verification of training quality in the training session of athletes. Most of the coaches in colleges and universities monitor the effect of physical training in real time by simply monitoring the changes in the physical indicators of the athletes after training as well as their experience, and there is a lack of scientific basis and effective feedback data to verify the effect of physical training of athletes in track and field. Due to the lack of scientific and effective feedback

monitoring data to understand the training status of athletes in a timely manner, the quality of training has not been effectively verified, making the physical training of athletes is not as effective as desired.
2) There is a lack of scientific and effective methods to assess the recovery effect of athletes in the recovery process. Nowadays, university sports education in the training of athletes tend to pay more attention to the ability to train, but ignored the fatigue and injury suffered by athletes in high-intensity training, unable to efficiently grasp the recovery of athletes, which to a certain extent affects the athletes' normal performance on the field and the development of their future careers.

3. Optimisation of training and recovery methods for university sports athletes

In order to address the challenges faced by higher education physical education in the training and recovery aspects of athlete development in the previous chapter, this paper will target the following training and recovery optimisation methods.
1) Training Optimisation Methods for Athlete Cultivation in Colleges and Universities. Using data mining technology, constructing athlete training quality monitoring system, realising timely and comprehensive collection of athletes' training and competition data, and adjusting athletes' training plan based on training data.
2) Recovery Optimisation Method for Athlete Cultivation in Colleges and Universities. Combined with data mining technology, we build an intelligent recovery assessment system for athletes, combining Kinect sensors to capture the key points of athletes' skeleton, and realise intelligent recovery state assessment on the basis of movement characteristics to formulate athletes' recovery plan.

Next, the above proposed optimisation method for athlete training and recovery is studied in depth.

3.1. Athlete Training Quality Monitoring System

3.1.1. Concept of multidimensional association rules:
1) Definition. Before defining multidimensional association rules, let us define unidimensional association rules [28]. Firstly, define 'dimension', in multidimensional database, each different predicate is called dimension.

A unidimensional rule is defined as follows.

An association rule that contains multiple occurrences of a single predicate (i.e., the predicate appears multiple times in the rule), i.e., a relationship between values of the same attribute. It is given in equation (1) below:

$$plays(X,"pingpang") => plays(X,"badminton") \tag{1}$$

Multidimensional rules are defined as follows.

An association rule that contains two or more predicates, i.e., relationships between values of different attributes. It is given in equation (2) below:

$$age(X,"20...29") \wedge sex(X,"male") => plays(X,"bad\min ton") \tag{2}$$

2) Classification. Interdimensional association rule. In the above equation (2) three predicates, "age", "gender" and "choice of sport" are included and each predicate appears only once in the rule. This is called having non-repeating predicates. An association-rule with non-repeating predicates is called an inter-dimensional association rule.
3) Mixed dimensional association rules. Example:

$$age(X,"20\cdots 29") \wedge plays(X,"pingpang") => plays(X,"bad\min ton") \tag{3}$$

As in equation (3) above, predicates such as "choose movement" appear twice in the rule, and this kind of association rule containing multiple occurrences of certain predicates is called a mixed-dimensional association rule.

3.1.2. Description of the Apriori Improvement Algorithm. The improved algorithm in this paper is used to mine multidimensional association rules using quantitative attribute discretisation. The improved algorithm is described in detail below.

1) Pre-organisation of data. The first step in data mining is to process the data. Multi-dimensional association rules mining before the need to analyze the attributes of the data table for processing, the processing method is in accordance with the requirements of the "quantitative attribute discretization", according to the actual significance of each value of the attribute hierarchical division, and with 0, 1, 2 ... and so on. It is important to note that there are no duplicate values between each level of the attribute.

The following is an example of badminton, where the scoring of the last beat is divided into points lost and points scored, which are represented by 0 and 1, respectively. The factors of scoring and losing points are divided into active scoring, active error, pressure error and non-pressure error, which are expressed by 0, 1, 2 and 3 respectively. The course of the ball's flight when struck is classified as straight, middle, opposite, or diagonal, and is denoted by 0, 1, 2, and 3, respectively, and so on. Along these lines, the domain of each property to be analysed is discretised into mutually exclusive levels.

Here two variables are defined, ShuNum, CengNum. ShuNum denotes the number of target attributes to be mined by the association rule, i.e., the number of multiple attributes to be analysed that are selected in the data table. cengNum denotes the number of values contained in the attribute with the highest value distribution among all the attributes, i.e., the maximum number of hierarchies. It should be emphasised here that the values of ShuNum and CengNum are not arbitrarily large, and considering the feasibility of the algorithm, less than 10 is suitable for implementation.

2) Scanning the database. One of the advantages of the improved multidimensional association rule algorithm is that during the implementation of the algorithm, the database is scanned only once, i.e., only one database read operation is carried out, which has a great advantage in terms of workload compared with other algorithms that scan the database several times, especially when the database is large enough to be millions of millions of times, the improved algorithm improves the algorithm's efficiency in a very considerable way.

The algorithm scans the database with a one-dimensional array Num[Range] to record specific hierarchical information, each attribute of each record to be stored, with the advancement of the scanning, in the encounter of the same hierarchical information is constantly added up, and finally obtained all the records of each attribute of each level of the corresponding count sum.

The initial value of Num[Range] is set to 0. After one scanning of the database, the number of records TotalNum and the one-dimensional array Num[Range] can be obtained, where the values in Num[Range] are the counts of different records with the same attributes and levels.

3) Finding frequent itemsets. With a one-dimensional array Num[Range] to store the various target elements and hierarchical information in the data table, the next step is to start looking for all the frequent itemsets.

Frequent 1-item set L_1. We assign ourselves the minimum support Sup, plus the number of records TotalNum output in the previous step, and use formula (4) to get SupNum, which represents the minimum number of Nums for the frequent item set:

$$SupNum = TotalNum * Sup \tag{4}$$

The final job of this step is to compare the size of the value of Num1[S, C] and the value of SupNum in turn, if Numl[S, C] < SupNum, no operation. If Num1[S, C] > SupNum, the corresponding S and C are the constituents of the frequent 1-item set, output L_1.

Frequent $n-$item sets L_n. The process of finding frequent 2-item sets L_2 is to scan Num[Range] starting from 0 to Max{Range} (which is similar to finding frequent 1-item sets), and if the value of Num[Range] $\neq 0$, one needs to look at C2 from the beginning, and if there is a record in C2 that matches the number on each bit of Range, which is set to be I, then add Num[Range] on top of the

counts in I Num[Range] is added. After checking Num[Range], the temporary data table C2 is to be finally cleaned up by deleting the rows with Num<SupNum, and the corresponding attributes and hierarchies in the remaining C2 are the constituents of the frequent 2-item set, output L_2.
4) Output the strong association rule. The above steps get the frequent item set $L_1, L_2, L_3 \cdots\cdots L_n$, and the following is to get the strong association rules from these frequent item sets. The minimum confidence level is set artificially and denoted by Con. To analyse the association rules between n attributes, 1 attribute is taken as the posterior term and the remaining $n-1$ attributes are taken as the anterior terms respectively, using the following equation (5):

$$Con = \frac{n\text{-}1 \text{ Number of attribute level combinations}}{n \text{ Number of parent attribute levels and}} \tag{5}$$

Among the obtained rules, the rules with confidence $\geqslant$ minimum confidence are outputted as the required strong association rules.

3.1.3. Algorithm evaluation. The improved algorithm performs only one read operation to the database and there are no duplicate values between the different levels of the same attribute for each record, thus producing fewer higher-order candidate sets than the Apriori algorithm [29].

The following is an example of generating a candidate 2-item set to calculate the difference between the two algorithms.

Let the number of frequent 1-item sets generated be n and the number of candidate 2-item sets obtained by concatenation be A_1. In the Apriori algorithm, A_1 is computed by equation (6):

$$A_1 = C_n^2 = \frac{n(n\text{-}1)}{2} \tag{6}$$

Let the number of attributes involved in the frequent 1-item set be n_1. For comparative simplicity, let the number of hierarchical levels of each attribute be similar, with a mean of n_2. The number of candidate 2-item sets produced by the connection is A_2. In the improved algorithm, A_2 is obtained by calculating through Eq. (7), Eq. (8):

$$n = n_1 \times n_2 \tag{7}$$

$$A_2 = C_{n_1}^2 \times C_{n_2}^1 \times C_{n_2}^1 = \frac{n(n\text{-}n_2)}{2} \tag{8}$$

Then the difference ΔA between the number of candidate 2-item sets produced by the Apriori algorithm and the improved algorithm is obtained by equation (9):

$$\Delta A = A_1 - A_2 = \frac{n(n_2\text{-}1)}{2} \tag{9}$$

It follows that $n > 0$, the number of attribute hierarchies in the improved algorithm must be at least 2, i.e., $n_2 > 1$, and hence $\Delta A > 0$, indicating that the candidate itemsets produced by the improved algorithm must be smaller than those produced by the traditional Apriori algorithm. The number of candidate 2-item sets in the improved algorithm is inversely related to the number of attribute hierarchies in the target database, i.e., as the number of hierarchies goes up, the number of 2-item sets goes down, and accordingly the overhead of the algorithm at the system level decreases accordingly.

3.1.4. System design. Combined with the improvement of Apriori algorithm based on multidimensional association rules proposed above, this section will design the athlete training quality monitoring system and propose the corresponding decision-making function regarding.
1) Data fusion processing. The data feature identification function of the sports training evaluation decision support system is:

$$P_c = \sum_{i=0}^{n}\sum_{j=0}^{n}\alpha(i,j)P(i,j) \tag{10}$$

Assuming that the starting code element for each of the above attributes is $C_0 = C_{N/2} = 0$, $C_{N-n} = C_n^*, n = 0,1,2,\cdots,N/2-1$, the model relationship between the sports training evaluation decision data and the cluster centre distribution is:

$$p_r = \frac{P_t}{(4\pi)^2\left(\frac{d}{\lambda}\right)^{\gamma}}\left[1+\alpha^2+2\varepsilon\cos\left(\frac{4\pi h^2}{d\lambda}\right)\right] \tag{11}$$

Combining the association criteria of sports training patterns, the feature identification of sports decision support system is performed based on the unused types of the obtained data. In this case, the category of association criterion attributes in the system is:

$$R_\beta X = U\{E \in U/R \mid c(E,X) \le \beta\} \tag{12}$$

$$R_\beta X = U\{E \in U/R \mid c(E,X) \le 1-\beta\} \tag{13}$$

For different data block types m_i and m_j, the iterative process of sports training decision making is obtained by using fuzzy mean method in the data types in conjunction with the association criterion of sports training pattern as:

$$\begin{aligned} S_b &= \sum_{i=1}^{e} p(\omega_i)(u_i-u)(u_i-u)^T \\ S_\omega &= \sum_{i=1}^{e} p(\omega_i)E\left[\frac{(u_i-u)(u_i-u)^T}{\omega_i}\right] \\ S_i &= S_b + S_\omega \end{aligned} \tag{14}$$

Where $p(\omega_i)$ is the set of association rule vectors for the sports training decision-making system, and the fusion of the information of the decision support system for the sports training model is realised according to the above computational equation.

In the use of data mining technology applied to the sports training mode decision-making system design discussion process, it is necessary to construct an association rule to tabulate the information of the data [30]. In the paper, adaptive weighting algorithm is used to construct the sports training pattern association criterion enlistment, and the final key rule mining weighting coefficient is obtained:

$$\omega_{sij}(n_0+1) = \omega_{sij}(n_0) - \eta_{sij}\frac{\partial J}{\partial \omega_{sij}} \tag{15}$$

A similarity analysis of the data associated in the decision support system for sports training models yields its adaptive learning process as:

$$\alpha_{desira}^{i} = \alpha_1 \cdot \frac{Density_i}{\sum_i Density_i} + \alpha_2 \frac{AP_i}{AP_{init}} \tag{16}$$

The associated data stored in the system is rearranged in the form of quintuples to obtain the probability density function for data mining as:

$$P_s = P_{2D}^{k}(1-P_{2D})^{N-1-k}\sum_{i=1}^{\infty}\lambda_s^i = \frac{\lambda_S}{1-\lambda_S} \tag{17}$$

Where λ_s, P_{2D} is the association dimension of the data, and the probability that the data can be effectively detected, respectively. The degree of dissimilarity between data clusters in the decision support system for sports training model is:

$$DisSim(A,B) = 1 - \left| \frac{SameDis(A) - SameDis(B)}{Dis(A) + Dis(B)} \right| \quad (18)$$

where $Dis(A), Dis(B)$ is the sample distribution density and sample semantic set of the database, respectively.
2) Specific design. In the paper, the improved Apriori algorithm frequent association rules as shown in Fig. 1 was chosen to carry out the design of the human-computer interaction system. Algorithm compilation is used to integrate the data mining technology with the program of the system to carry out the design of the sports training mode decision support system. The main functional modules of the system are communication module, programme module, database module and data output module. The Conti-ki bus technology is used to transmit and coordinate the data types of the sports training mode decision support system, and at the same time, the integrated digging system can be realised by combining the VIX integrated control technology. The data sensing of the decision support system is constructed according to the 6LoWPAN protocol stack. The design of the wireless sensing network system adopts Atmel1284P as the main chip to dig the overall IoT address allocation and mobilisation of the sports training mode decision-making system. When the task bar address is determined, the human-computer interaction of the system is realised through the TaskBasic interface program.

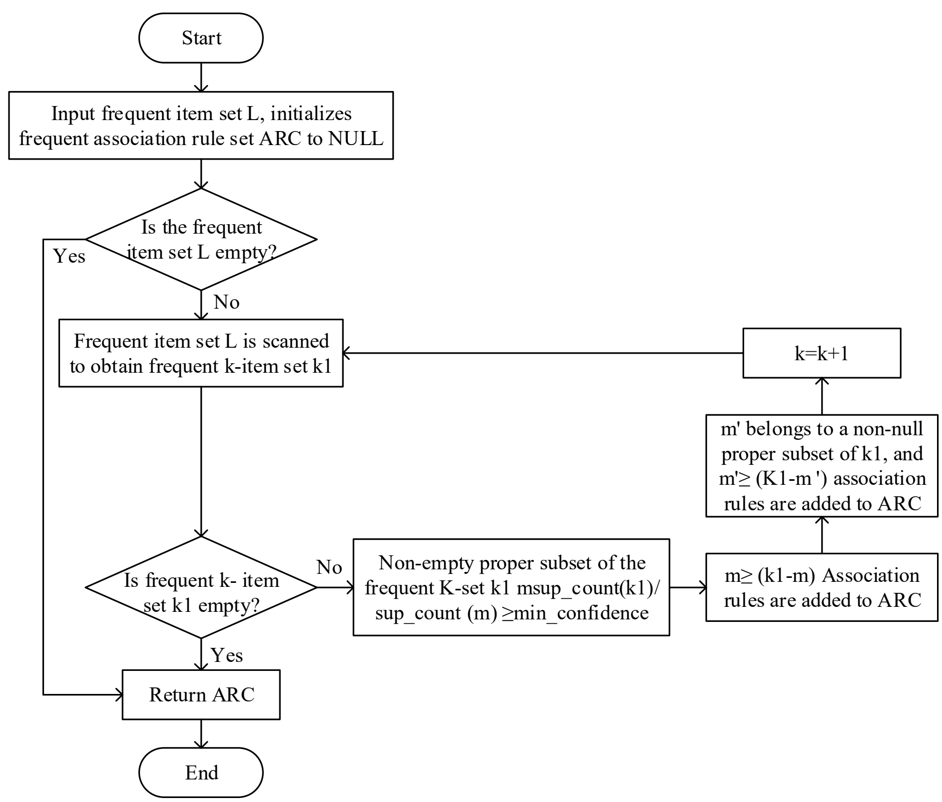


Fig. 1. Frequent correlation rule process

3.2. Athlete Intelligent Recovery Assessment System

3.2.1. Intelligent recovery assessment programme design:

1) Movement node data collection. The designed administration process of this paper is a trial in which each subject is required to complete a prescribed paradigm movement. A single straight line walk of

10m, walking 5 times in one trial, and repeating 5 trials, the skeletal keypoints captured by the Kinect sensor are used. During the trial, the 3D coordinate positions of the subjects' skeletal nodes are captured and stored in real time.
2) Deriving gait characteristics from digital human motion models. In this paper, the time for one lower limb of the subject to complete the gait cycle from the time the foot hits the ground to the time it hits the ground again is taken as a gait cycle to extract the test data. A complete gait cycle is divided into 100 intervals, according to which the skeletal keypoint detection data signal is resampled 100 times. The ground reaction force data was downsampled based on the low-frequency skeletal critical point detection data, and then the ground reaction force data was resampled 100 times over a complete gait cycle using nearest neighbour interpolation. The amplitude of the ground reaction force data was also normalised by body weight.

3.2.2. Filtering the optimal subset of features. In this paper, after extracting the features of the gait pattern, a fuzzy binary comparison decision-making method based on information gain is used to rank the importance of the influence factors of the gait pattern.Firstly, the binary comparison between two and two features is achieved by calculating the information gain of all the features so as to establish the binary comparison level. Then, using the fuzzy relative comparison function, a fuzzy phase sum matrix is established, and all features are ranked by determining the $\lambda -$ intercept matrix. Finally, the SFS strategy is used to add the features with the highest importance to the feature subset one by one and train the classification model so as to derive the optimal feature subset.
1) Calculate the binary comparison level between features. Getting the binary comparison level between features is the key to fuzzy relative comparison decision-making, usually the binary comparison of features in the fuzzy decision-making method adopts the expert scoring method, in view of the reliability of the qualitative analysis method is not high, this paper uses the reduction of information gain (RED) to achieve the binary comparison between the two features. the higher the RED value, that is, the greater the impact of this feature on the prediction of the sample results, which indicates that the degree of its importance is The higher the RED value, the greater the impact of the feature on the prediction results of the sample, and thus the higher its importance.

There is an input feature n , and after permutation, all features are divided into two-two combinations of feature subset $\{(x_b x_j) \mid i, j \in n, i \neq j\}$, i.e., there are $\frac{1}{2}C_n^2$ feature subsets. In order to evaluate the relative importance of feature x_i and feature x_j , the procedure is as follows.

Step 1, compute x_i the information gain $g(x_{i,D})$ and x_j the information gain $g(x_j, D)$ on the sample data $D=\{(x_i,x_j)\}$ containing only two features.

Step 2, add random noise to x_i in the data set, with the sample values of x_j unchanged, and use the perturbed sample data to obtain a new data set $\underline{D}=\{(\underline{x_i},x_j)\}$. Recalculate the information gain $g(\underline{x_i},D)$ of x_i .

Step 3, repeat step 2 N times to obtain the RED_{xi} of feature xi by taking the difference mean of $g(xi,D)$ and $g(\underline{x_i},\underline{D})$. The importance of feature xi is calculated as:

$$RED_{x_i} = \frac{1}{N}\sum_{i}^{n}(g(xi,D)-g(\underline{xi},\underline{D})) \tag{19}$$

Step 4, add noise randomly to x_j in the dataset and leave the sample values of x_i unchanged to obtain a new dataset $D=(x,x,y)$ using the perturbed sample data. Recalculate the information gain $g(x_{jD})$ for x_i . Repeat the process N times and then use Eq:

$$RED_{x_j} = \frac{1}{N}(g(xj,D)-g(\underline{xj},\underline{D})) \tag{20}$$

Solve for RED_{xj} . The binary comparison level between feature x_i and feature x_j is then obtained (RED_{xi}, RED_{xj}).

Step 5, repeat steps 1-4 to get the binary comparison level between all features.

2) Obtain the optimal subset. The SFS strategy is used to add features of high importance one by one to the feature subset and train the classification model. When the recognition accuracy is maximum, the corresponding number k of input variables is the dimension of the optimal feature subset. Thus, the top k variables with maximum importance can be selected as the optimal input feature set for the model.

In this paper, support vector machine, random forest, and neural network are used to construct the training model [31]. Let the training sample set $(x_j, y_j), j = 1, 2, \ldots, n; x_j \in R^U$ be the input variables. $y_j \in R$ is the output variable, SVM maps the input space to the high-dimensional feature space through the nonlinear function $\phi(\cdot)$, and constructs the linear regression function in this space:

$$\begin{aligned} &\max_{\alpha,\hat{\alpha}} \sum_{i=1}^{n} y_i(\hat{\alpha}_i - \alpha_i) - \varepsilon \sum_{i=1}^{n} y_i(\alpha_i + \hat{\alpha}_i) \\ &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j) y_i y_j \kappa(x_i, x_j) \\ &s.t. \sum_{j=1}^{n}(\hat{\alpha}_i - \alpha_i) = 0, 0 \le \alpha_i; \hat{\alpha}_i \le C \end{aligned} \tag{21}$$

where α_i and α_j are the Lagrange multipliers and $\kappa(x_i, x_j)$ is the kernel function that satisfies Mercer's condition, in this study the radial basis kernel function is used, i.e.:

$$\kappa_{RBF}(x_i, x_j) = \exp(-\| x_i - x_j \|)^2 / 2\sigma^2, \sigma > 0 \tag{22}$$

The random forest algorithm is a combinatorial model consisting of a set of categorical decision subtrees $\{h(x, \theta_t), t = l, 2, ..., T\}$. Where θ_t is a random variable obeying an independent homogeneous distribution, x denotes the independent variable, and T denotes the number of decision trees. The idea of integrated learning is used to take the mean value of each decision subtree $\{h(x, \theta_t)\}$ as the prediction result:

$$\bar{h}(x) = \frac{1}{T}\sum_{t=1}^{T}\{h(x, \theta_t)\} \tag{23}$$

where $h(x, \theta_t)$ is the output based on x and θ.

3.2.3. Intelligent recovery assessment. Random forest algorithm has the advantages of not easy to overfitting, fast training speed, and suitable for dealing with high-dimensional data, because the random forest algorithm itself can not optimise its parameters, this paper introduces and improves the whale algorithm to achieve the automatic optimisation of random forest model parameters.

Firstly, a chaotic transformation is used to initialise the solution position, compared with random initialisation, the solution distribution generated by the chaotic transformation is more uniform and stable, more diverse, which is conducive to the global convergence of the algorithm and can effectively avoid falling into the local optimum. Then the random walk strategy is introduced to obtain the improved whale algorithm. The random walk strategy can effectively prevent the algorithm from falling into the local optimum too early and increase the exploration ability of the algorithm. The mathematical expression is shown below:

$$\tilde{X}_{i+1} = \begin{cases} X_{i+1} + (X_i^*(t) - AD) \times RW(l) & p < 0.5 \\ X_{i+1} + (D'e^{bl}\cos(2\pi t) + X_i^*(t)) \times RW(l) & p \ge 0.5 \end{cases} \tag{24}$$

$$RW(l) \approx l^{-(1+\vartheta)} \tag{25}$$

$$l = \frac{\alpha}{|\beta|^{1/\alpha}}, \alpha, \beta \sim N(0, \delta^2) \tag{26}$$

$$\delta^2 = \left[\frac{\Gamma(\vartheta)}{\vartheta \Gamma((1+\vartheta)/2)} \frac{\sin(\pi \vartheta / 2)}{2^{(1+\vartheta)/2}} \right]^{2/\vartheta} \tag{27}$$

where l denotes the step size, $\Gamma(.)$ denotes the gamma function, and $\vartheta = 1.5$.

4. Optimising practices for athlete training and recovery

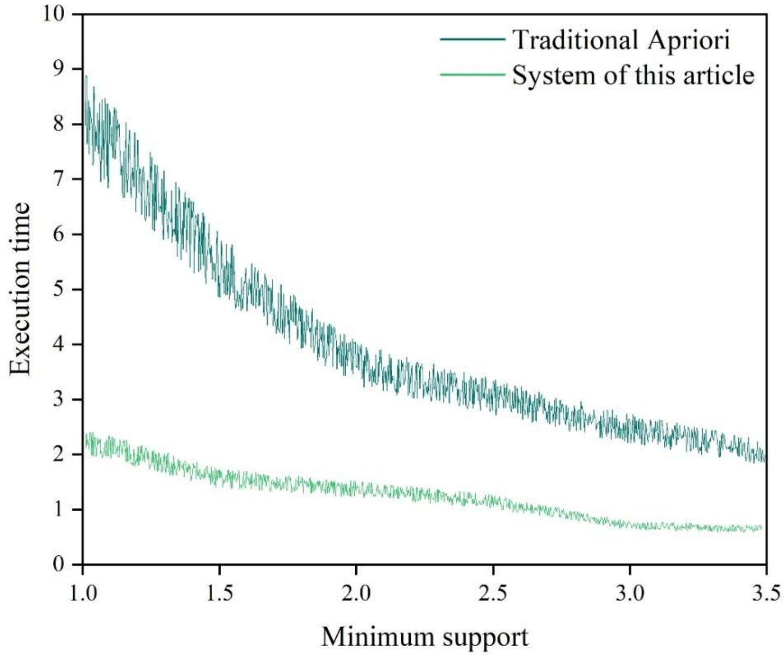
In order to verify the validity and applicability of the training and recovery methods for athletes proposed in this paper, this paper will adopt the method of before and after experiment comparison, and 20 track and field athletes from University H of Jiangxi Province will be selected as the experimental subjects. Before the experiment, the 20 track and field athletes will be tested for physiological and biochemical indexes. The study will last for 12 weeks, totalling 3 months, and the experimental location is the stadium on the campus of University H.

During the experiment, the athlete training quality monitoring system and intelligent recovery assessment system constructed in this paper will be used as a sports teaching tool, and the training and recovery optimisation method proposed in this paper will be combined to formulate the athlete training and recovery planning scheme.

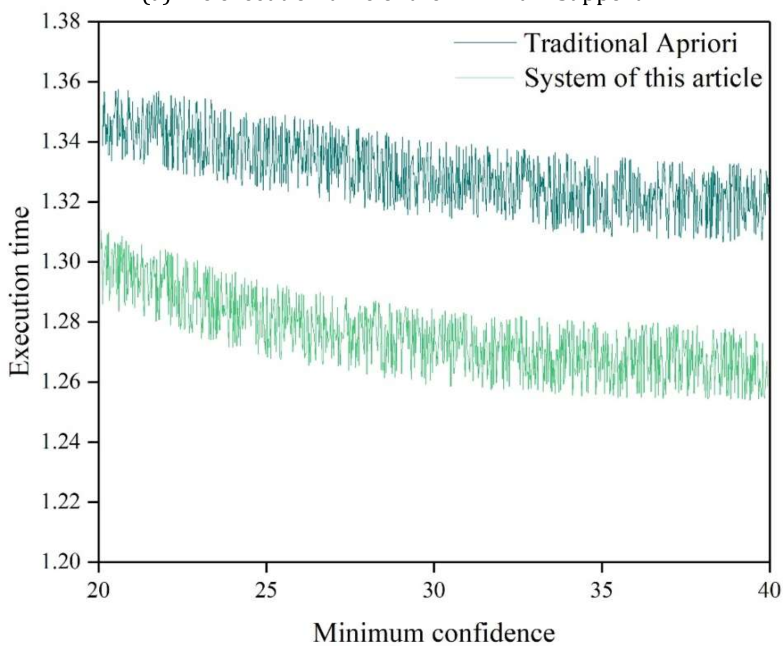
4.1. Testing of Optimisation Methods for Athlete Training and Recovery

Before formally carrying out the practice of athlete training and recovery optimisation, the training and optimisation methods of university athlete training, i.e., the athlete training quality monitoring system and the intelligent recovery assessment system proposed in this paper, will be functionally tested. The data used for the test are the relevant sports data of 20 track and field athletes in University H.

4.1.1. Testing of systems for monitoring the quality of athlete training. It is known that when constructing the athlete training quality monitoring system, this paper adopts the multidimensional association rules to improve the Apriori algorithm, and this section will take the traditional Apriori algorithm as a comparison to explore the changes in the execution time of the athlete training quality monitoring system constructed in this paper with the increase of the minimum support degree and the minimum confidence degree, as shown in Figure 2. As can be seen from the figure, combined with the improved Apriori algorithm based on multidimensional association rules, the system in this paper has a smaller reaction time in the case of minimum support. When the minimum support is 1, the execution time of this paper's system is 2.48s, while the traditional Apriori algorithm is 9s, and the difference reaches the maximum of 6.52s, and with the growth of the minimum support, the difference between the two is also shrinking. Similarly, when the minimum confidence is small, the system of this paper combined with the improved Apriori algorithm performs better in terms of execution time. When the minimum confidence is the smallest 20, the execution time of this paper's system is 1.31 s, which is higher than that of traditional Apriori algorithm 0.05 s. When the minimum confidence grows, the gap of execution time is also shrinking, but the execution time of this paper's system is always lower than that of traditional Apriori algorithm. Overall, the athlete training quality monitoring system constructed in this paper has good execution efficiency.



(a)The execution time of the minimum support



(b)Execution time under minimum confidence

Fig. 2. The change in execution time

4.1.2. Intelligent recovery assessment system testing. In the intelligent recovery assessment system for athletes constructed in this paper, Kinect sensors are used to capture the key points of the athlete's skeleton and derive the gait characteristics of the athlete from the digital human model. In this section, the athlete's joint data obtained from the system of this paper will be compared with the athlete's real joint data to verify its accuracy. The absolute and relative errors between the athletes' joint data obtained by applying the system of this paper and the real joint data of track and field athletes are shown in Figure 3. As can be seen from the figure, the absolute error between the joint data of track and field athletes obtained by the system of this paper and the real data is less than 10px, and the average value of the absolute error is 4.87px. The relative error between the two sets of data is only 5.01% at the highest, and the average value of the relative error is 3.67%. Obviously, the intelligent recovery assessment system for athletes constructed in this paper has high accuracy, and the extraction results of its athletes' joint data are close to the real data.

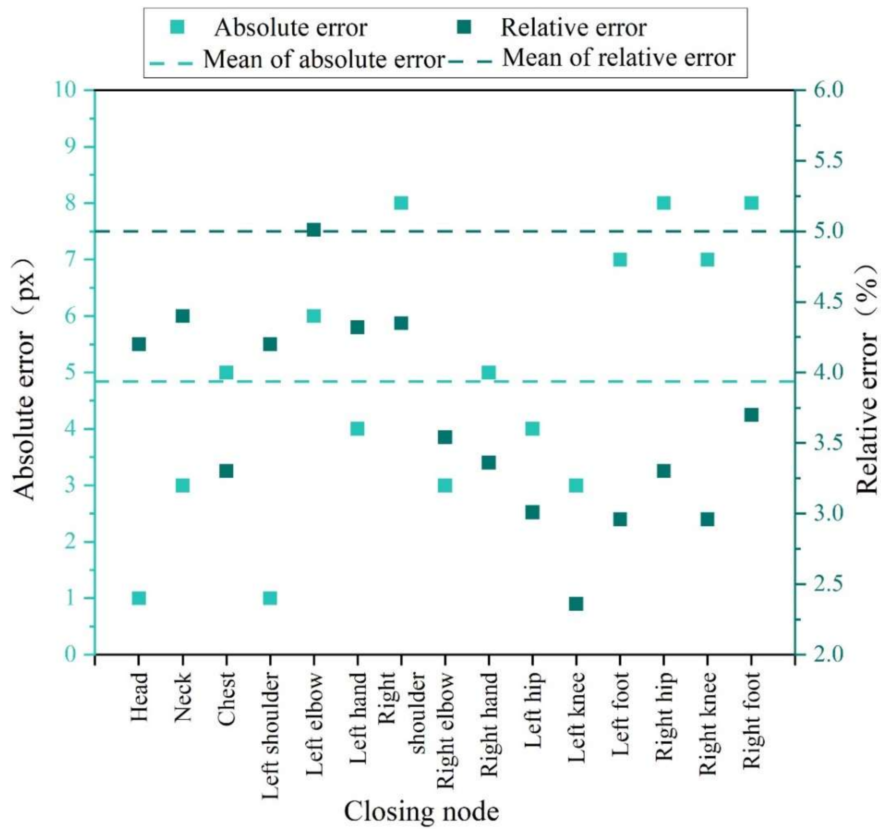


Fig. 3. Data contrast

4.2. Analysis of Results of Optimised Athlete Training and Recovery Practices

4.2.1. Analysis of the effects of training optimization. At the end of the 12-week training and recovery optimisation experiment, a fitness test was carried out on 20 track and field athletes to understand the athletes' physical fitness and test the results of this training experiment. The physical fitness test consisted of 7 items, including over-the-top squat, hurdle step, linear lunge squat, shoulder joint flexibility, active straight knee raise, stability push-up, and trunk rotation test, which were referred to by S1~S7 in order. The comparison of the physical fitness test scores of track and field athletes before and after the training experiment is specifically shown in Figure 4, where A and B refer to before and after the experiment, respectively. As can be seen from the figure, except for the shoulder joint flexibility index, 20 track and field athletes of University H showed significant differences in the $P<0.05$ of the six indexes of over-the-top deep squat, hurdle step, linear lunge squat, active straight knee raise, stability push-ups, and trunk rotation test before and after the experiment, and the value-added of the scores before and after the training experiments were 0.12, 0.23, 0.13, 0.16, and 0.25, respectively, 0.67. The total score of fitness test of track and field athletes increased from 18.19 to 19.8 before the experiment, and the optimisation method of athletes' training proposed in this paper contributes to the improvement of athletes' training quality.

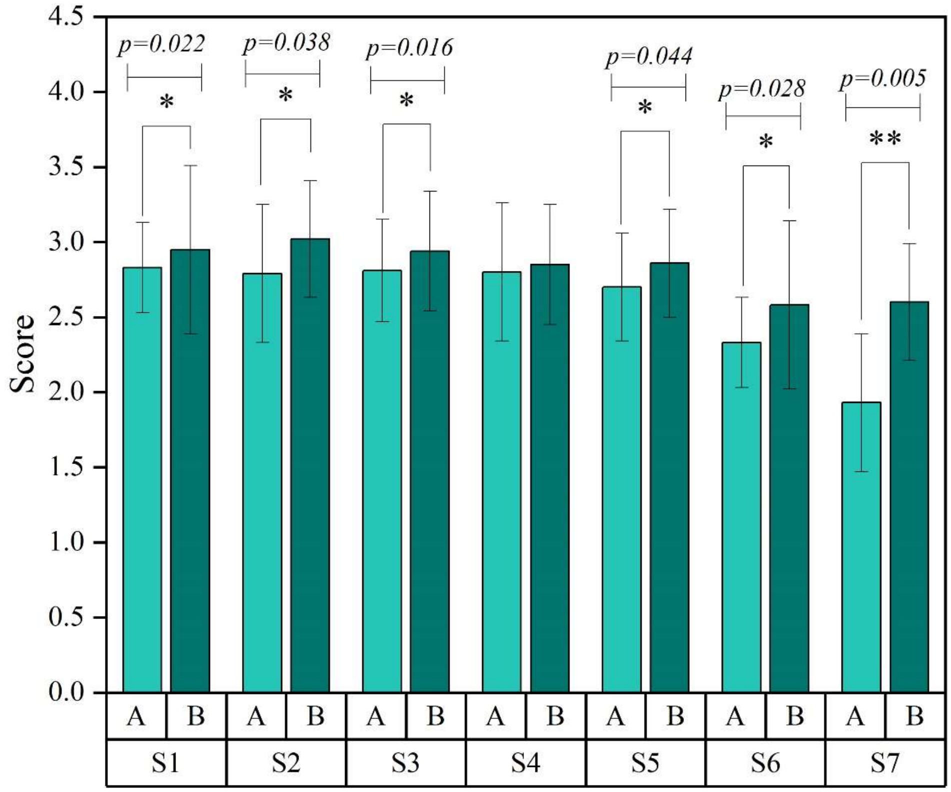


Fig. 4. Physical testing data

4.2.2. Recovery Optimisation Effectiveness Analysis. In this athlete recovery optimisation experiment, four body health data, heart rate, blood lactate, serum creatine kinase, and haemoglobin, will be recorded uniformly for 20 track and field athletes in a calm state before training. The same body data measurements will be taken from the track and field athletes when they are in a fatigued state at the end of training and when they are in a recovered state after exercise recovery using the athlete recovery optimisation method proposed in this paper. The health indicator data of track and field athletes at different time stages are specifically shown in Figure 5.

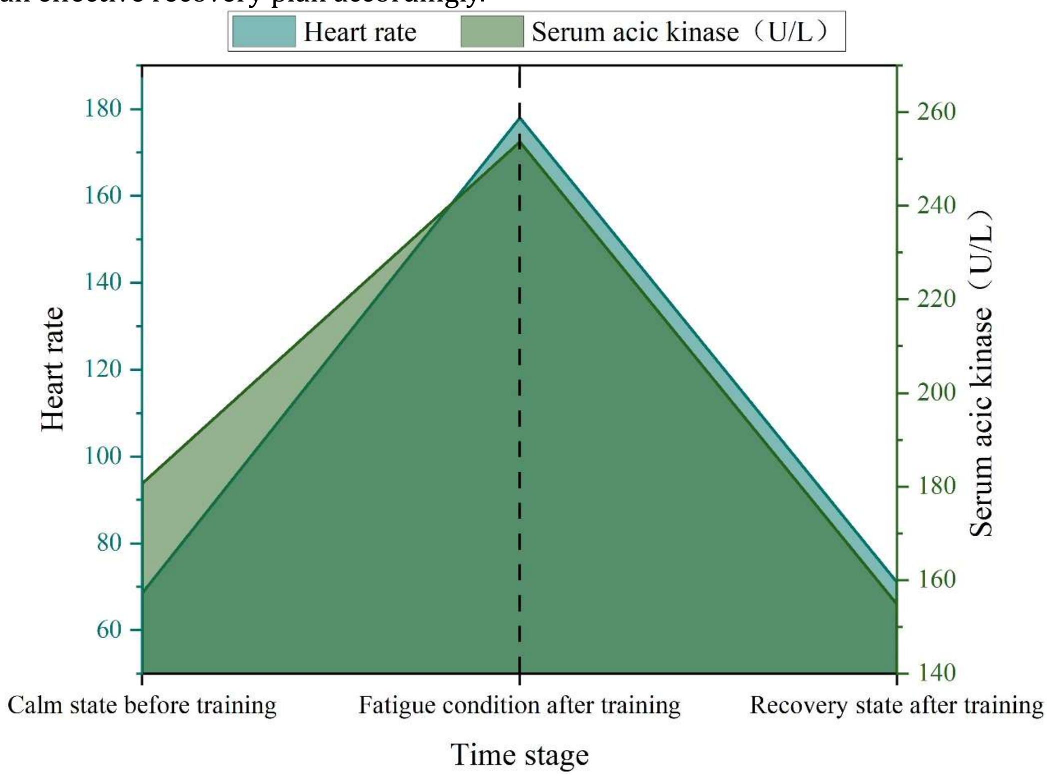
Heart rate is one of the common indicators reflecting the cardiac load during muscle activity. From the statistics in the figure, it can be found that the heart rate of track and field athletes is the highest when they are in a fatigue state after performing sports training, up to 178 beats. After using the optimisation method of athlete recovery in this paper, the heart rate of track and field athletes decreased to 71 times, which basically restored to the level of heart rate in the calm state before training.

The value of blood acid accumulation is one of the main indicators for evaluating physical fatigue. A two-by-two comparison of the blood acid accumulation values of the athletes at the three time stages showed that the blood lactate values in the fatigue state after exercise training (5.87 mmol/L) and after recovery from exercise (2.14 mmol/L) were increased compared with those in the calm state before training (1.98 mmol/L), but the blood lactate values after recovery from exercise had a significant decreased and were only slightly higher than blood lactate values in the untrained calm state.

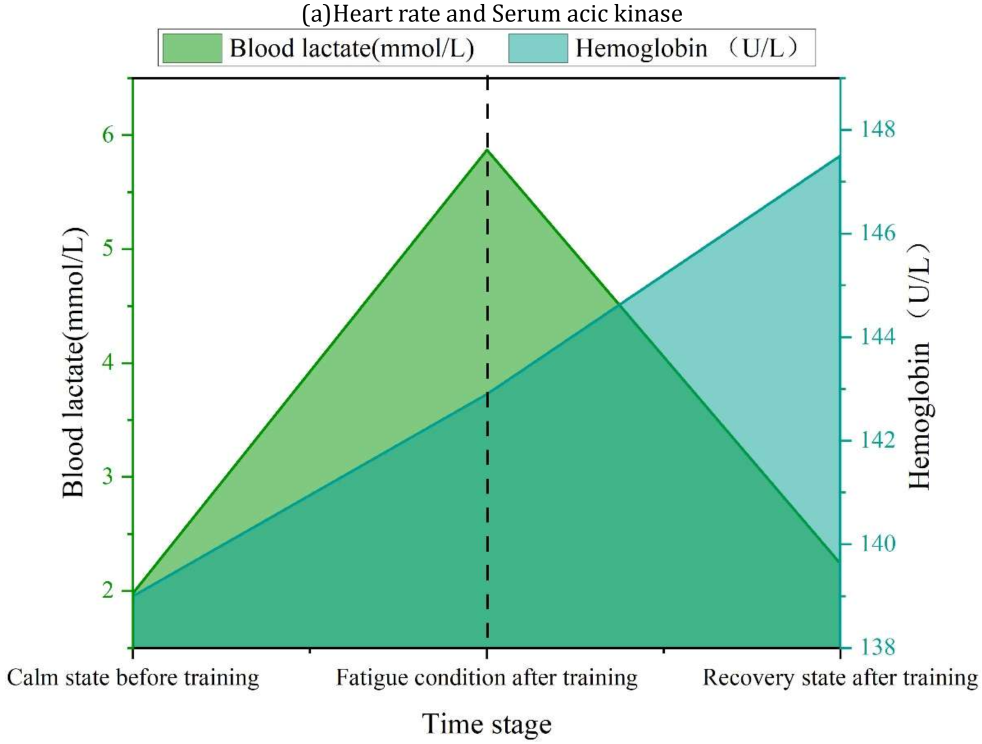
Creatine kinase is one of the most important indicators of fatigue recovery from exercise training and muscle strain. As can be seen from the graph, serum creatine kinase was highest in the fatigued state of athletes after sports training, reaching 253.7 U/L, and lowest after using the method of recovery optimisation of athletes in this article, with serum creatine kinase decreasing to 154.8 U/L, even lower than the serum creatine kinase level in the calm state before training (180.6 U/L).

The role of haemoglobin in the body is to carry and transport oxygen, and it is an important indicator of aerobic metabolism in athletes. From the data, it can be seen that after using the optimisation method of athlete recovery in this paper, the haemoglobin content of track and field athletes was significantly improved to 147.5 U/L, which is higher than the haemoglobin content of track and field athletes in the state of calmness in order to carry out the training (139 U/L) and in the state of fatigue after sports training (142.9 U/L).

The analysis of the above health index data shows that the proposed recovery optimisation method can effectively alleviate the fatigue state of athletes after training, and the constructed intelligent recovery assessment system for athletes can effectively assess the recovery state of athletes and formulate an effective recovery plan accordingly.



(a)Heart rate and Serum acic kinase



(b)Blood lactate and Hemoglobin

Fig. 5. Data on health indicators

4.2.3. Analysis of satisfaction with training and recovery programmes. In this section, data on the satisfaction of 20 track and field athletes at the University of H with this training and recovery planning programme will be collected at the end of the training and recovery experiment. Each satisfaction dimension and the specific development of satisfaction indicators under the dimension are shown in Table 1.

Table 1. The satisfaction of training and recovery plan

Dimension	Index
Training condition dimension	Method adjustment
	Fatigue recovery
	State adjustment
Coach factor dimension	Training feedback
	Training guidance
Personal dimension	Training completion
	Pay and return

The satisfaction index scores under each satisfaction learning are specifically shown in Figure 6, where the mean value is the average value of each dimension, and the maximum and minimum values are the highest and lowest satisfaction scores among the 20 track and field athletes' scores. The mean values of training status, coaching factors and personal situation dimensions were 4.35, 4.425 and 4.38 respectively, and under the training status dimension, the ranking of each satisfaction indicator according to the level of scores was Method Adjustment (4.51) > Condition Adjustment (4.37) > Fatigue Recovery (4.16), of which the mean value of Fatigue Recovery indicator was lower than the overall average of dimensions (4.35), reflecting the fact that the internal satisfaction of track and field athletes was lower than the overall average of dimensions (4.35).), reflecting the large fluctuations in score differences within track and field athletes. Satisfaction indicators with scores lower than the dimension mean also included the training feedback indicator in the coaching factors dimension (4.25<4.425) and training completion in the personal situation dimension (4.35<4.38). Athletes' opinions on improvement can be collected to further optimise the training methods. Overall, the mean value of each satisfaction indicator was greater than 4. The track and field athletes who participated in this training and recovery optimisation experiment were more satisfied with the training and recovery optimisation methods during the experiment as a whole.

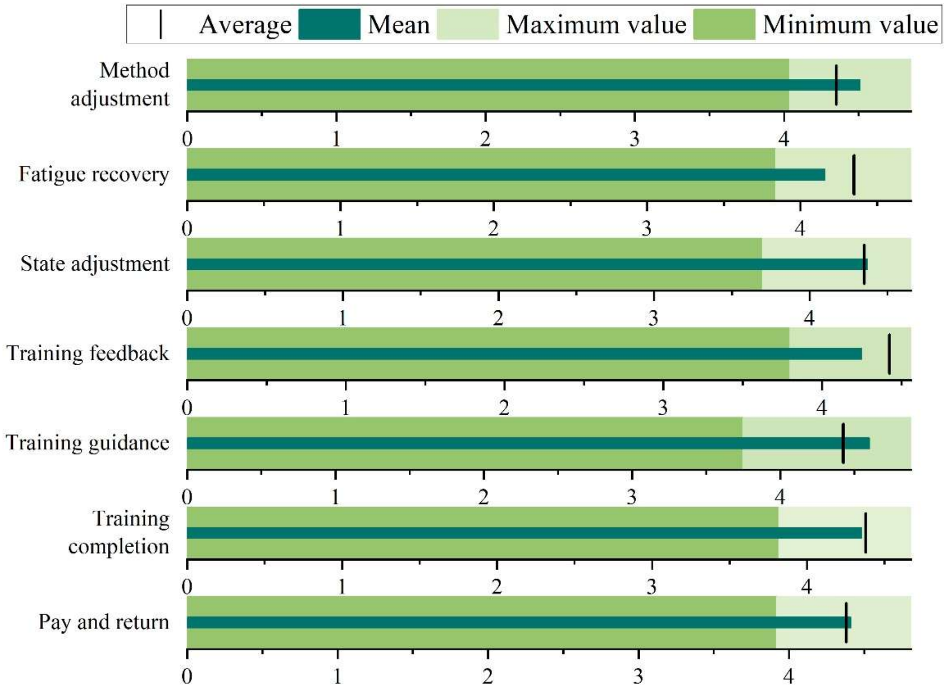


Fig. 6. Data of satisfaction of training and recovery plan

5. Conclusion

Based on data mining technology, this paper proposes a training and recovery optimisation method for college sports athletes, and builds up an athlete training quality monitoring system and an intelligent recovery assessment system. Twenty track and field athletes from University H in Jiangxi Province were selected as experimental objects to verify the effectiveness and applicability of the training and recovery optimisation method of this paper.

Using the data of 20 track and field athletes from University H as test data, the performance test of the athlete training quality monitoring system and intelligent recovery assessment system constructed in this paper was carried out. When the minimum support and minimum confidence are 1 and 20, the difference between the training quality monitoring system of this paper and the traditional Apriori algorithm is 6.52 s and 1.31 s. With the increase of the minimum support and minimum confidence, the execution time of the training quality monitoring system of this paper is always lower than that of the traditional Apriori algorithm. The average value of relative error between the athlete's joint data and the real joint data obtained from the intelligent recovery evaluation system constructed in this paper is 3.67%, and the average value of absolute error is 4.87px.

Through the practice of training and recovery optimisation, the pre- and post-experiment data of six indexes of over-the-top squat, hurdle step, linear lunge squat, active straight knee raise, stability push-up, and trunk rotation test of the track and field athletes in the physical fitness test showed significant differences ($P<0.05$), and the total score of the physical fitness test was improved from 18.19 to 19.8 before the experiment.In terms of the physical fitness data, after adopting the athlete recovery optimisation method, the athlete's heart rate decreased to 71 beats, the blood lactate value was 2.14 mmol/L, the serum creatine kinase decreased to 154.8 U/L, and the haemoglobin level was elevated to 147.5 U/L, effectively alleviating the athlete's fatigue state after training. For the training and recovery programmes developed in practice, the mean values of the dimensions of training status, coaching factors, and personal situation of the track and field athletes were 4.35, 4.425, and 4.38, respectively, which were all greater than 4, and the overall attitude was satisfactory.

References

[1] Charlina, C., Roziah, R., Ismail, S., Piliang, W. S. H., Siswanto, S., Setyawan, H., ... & Tafuri, F. (2024). Effective Verbal Communication in Physical Education Learning and Sports Coaching to Improve Achievement and Health: A Systematic Review. Retos, 56, 1139-1147.

[2] Oripjonova, R., & Tuychiyev, A. (2022). THEORETICAL FOUNDATIONS OF PHYSICAL EDUCATION AND SPORTS TRAINING IN WOMEN'S HEALTH PROMOTION. THE ROLE OF SCIENCE AND INNOVATION IN THE MODERN WORLD, 1(1), 106-110.

[3] Feu, S., García-Rubio, J., Gamero, M. D. G., & Ibáñez, S. J. (2019). Task planning for sports learning by physical education teachers in the pre-service phase. PloS one, 14(3), e0212833.

[4] Gümüs, H., Gençoglu, C., & Sahin, T. (2020). Physical Education and Sports: Bibliometric Analysis of the ERIC Database. International Online Journal of Education and Teaching, 7(4), 1823-1837.

[5] Kozin, S., Kozina, Z., Korobeinik, V., Cieślicka, M., Muszkieta, R., Ryepko, O., ... & Bejtka, M. (2021). Neuromuscular training for injury prevention of students-rock climbers studying in the specialty" Physical Education and Sports": a randomized study. Journal of Physical Education and Sport, 21, 1251-1259.

[6] Romanchuk, S., Anokhin, E., Tychyna, I., Dobrovolskii, V., Pidleteichuk, R., Homanyuk, S., ... & Klymovych, V. (2020). The impact of mass sports work in educational institution on the formation of cadets' value attitude towards physical education. Sport Mont, 18(1), 81-86.

[7] Byshevets, N., Iakovenko, O., Stepanenko, O., Serhiyenko, K., Yukhno, Y., Goncharova, N., ... & Blystiv, T. (2021). Formation of the Knowledge and Skills to Apply Non-Parametric Methods of Data Analysis in Future Specialists of Physical Education and Sports. Sport Mont, 19.

[8] Gennadievich, B. A. (2020). Machine learning and data mining activity results when using projectiles in different sports. International Journal, 9(3).

[9] Yuan, C., Yang, Y., & Liu, Y. (2021). Sports decision-making model based on data mining and neural network. Neural Computing and Applications, 33, 3911-3924.

[10] Dong, Y., & Dong, J. (2024). Evaluation and Analysis of Decision Support for College Fitness Teaching Under Embedded Data Mining Training. International Journal of High Speed Electronics and Systems, 2440085.

[11] Pan, L. (2019). A big data-based data mining tool for physical education and technical and tactical analysis. International Journal of Emerging Technologies in Learning (Online), 14(22), 220.

[12] Wang, X., Huang, D., & Zhao, X. (2020). Design of the Sports Training Decision Support System Based on the Improved Association Rule, the Apriori Algorithm. Intelligent Automation & Soft Computing, 26(4).

[13] Sands, W. A., Kavanaugh, A. A., Murray, S. R., McNeal, J. R., & Jemni, M. (2017). Modern techniques and technologies applied to training and performance monitoring. International journal of sports physiology and performance, 12(s2), S2-63.

[14] Xie, J., Chen, G., & Liu, S. (2021). Intelligent badminton training robot in athlete injury prevention under machine learning. Frontiers in neurorobotics, 15, 621196.

[15] Phatak, A. A., Wieland, F. G., Vempala, K., Volkmar, F., & Memmert, D. (2021). Artificial intelligence based body sensor network framework—narrative review: proposing an end-to-end framework using wearable sensors, real-time location systems and artificial intelligence/machine learning algorithms for data collection, data mining and knowledge discovery in sports and healthcare. Sports Medicine-Open, 7(1), 79.

[16] Li, C., & Li, Y. (2020). Feasibility analysis of vr technology in physical education and sports training. IEEE Access.

[17] Karimov, F. X. (2022). Scientific and Methodological Bases of Normalization of Loads in Physical Education and Mass Sports-Health Training. European Journal Of Business Startups And Open Society, 2(2), 24-28.

[18] Opstoel, K., Chapelle, L., Prins, F. J., De Meester, A., Haerens, L., Van Tartwijk, J., & De Martelaer, K. (2020). Personal and social development in physical education and sports: A review study. European Physical Education Review, 26(4), 797-813.

[19] Cojocaru, A. M., Cojocaru, M., Jianu, A., Bucea-Manea-Țoniș, R., Păun, D. G., & Ivan, P. (2022). The impact of agile management and technology in teaching and practicing physical education and sports. Sustainability, 14(3), 1237.

[20] Silva, A., Ferraz, R., Forte, P., Teixeira, J. E., Branquinho, L., & Marinho, D. A. (2022). Multivariate training programs during physical education classes in school context: theoretical considerations and future perspectives. Sports, 10(6), 89.

[21] Sgrò, F., Bracco, S., Pignato, S., & Lipoma, M. (2018). Small-sided games and technical skills in soccer training: Systematic review and implications for sport and physical education practitioners. Journal of sports science, 6(1), 9-19.

[22] 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.

[23] Teixeira, J. E., Encarnação, S., Branquinho, L., Morgans, R., Afonso, P., Rocha, J., ... & Forte, P. (2024). Data Mining Paths for Standard Weekly Training Load in Sub-Elite Young Football Players: A Machine Learning Approach. Journal of Functional Morphology and Kinesiology, 9(3), 114.

[24] Yin, Z., & Cui, W. (2021). Outlier data mining model for sports data analysis. Journal of Intelligent & Fuzzy Systems, 40(2), 2733-2742.

[25] Zhang, S., & Mao, H. (2021). Optimization analysis of tennis players' physical fitness index based on data mining and mobile computing. Wireless Communications and Mobile Computing, 2021(1), 9838477.

[26] Naglah, A., Khalifa, F., Mahmoud, A., Ghazal, M., Jones, P., Murray, T., ... & El-Baz, A. (2018, December). Athlete-customized injury prediction using training load statistical records and machine learning. In 2018 IEEE international symposium on signal processing and information technology (ISSPIT) (pp. 459-464). IEEE.

[27] Rajšp, A., & Fister Jr, I. (2020). A systematic literature review of intelligent data analysis methods for smart sport training. Applied Sciences, 10(9), 3013.

[28] Liu Yafei,Li Jiye,Ren Zhaoxu & Li Jun. (2022). Research on Personalized Recommendation of Higher Education Resources Based on Multidimensional Association Rules. Wireless Communications and Mobile Computing.

[29] Zheng Tang,Zhengwei Jiang,Ying Li,Haowei Yuan,Jiayu Han & Chao Chen. (2024). Research on the Association Analysis of Online Learning Behaviors Based on the Apriori Algorithm. Frontiers in Computing and Intelligent Systems(2),18-22.

[30] Wang Hua. (2024). An Empirical Study of Data Mining Technology in English Learning Outcome Prediction. International Journal of e-Collaboration (IJeC)(1),1-14.

[31] Maçãs Catarina,Campos João R,Lourenço Nuno & Machado Penousal. (2024). Visualisation of Random Forest classification. Information Visualization(4),312-327.