SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(1), 1071.
https://doi.org/10.62617/mcb1071

Article

Exercise training load prediction based on improved genetic algorithm

Bin Liu[1], Junhong Chen[2,*], Qian Wang[1], Wenwen Hu[1]

[1] Department of Physical Education, Institute of Disaster Prevention, Sanhe 065201, China
[2] College of Humanities, Hebei Oriental University, Langfang 065001, China
*** Corresponding author:** Junhong Chen, chenjunhong202304@163.com

CITATION

Liu B, Chen J, Wang Q, Hu W. Exercise training load prediction based on improved genetic algorithm. Molecular & Cellular Biomechanics. 2025; 22(1): 1071. https://doi.org/10.62617/mcb1071

ARTICLE INFO

Received: 10 December 2024
Accepted: 27 December 2024
Available online: 14 January 2025

Abstract: With the advancement of digital information age, all kinds of sports also widely used digital mining technology, the technology can promote athletes' professional level and physical quality, is in the field of sports improve athletes and coaches more effective way, but the bicycle sports for digital mining and research is still in the early state. Cyclists load ability in training is mainly influenced by the body function, the relationship between the two is complicated, this paper is the training load ability and physical function as a research object, explore the specific connection between the two aspects of cyclists, by constructing the cyclist training load prediction model, to apply in the "bicycle team training analysis system". Strive to promote the overall development of cycling, to provide a basis for the scientific training of talents and training. The cyclist training load prediction model created in this paper can provide a scientific design basis for the training of athletes. When analyzing the body function, this paper mainly analyzes the influence of athletes on conforming sports through 25 aspects such as maximum oxygen intake, functional threshold power and blood oxygen saturation. Because the factors of athletes and the predicted results show non-linear correlation, this paper selects the BP neural network algorithm in the model algorithm, and the adaptive genetic algorithm of the selection operator is adjusted and optimized to clarify the initial weights and thresholds of BP neural network. Finally, the practice has proved that the improved algorithm has a more prominent global optimization ability, and the accuracy of the model has reached 93.28%.

Keywords: ethnic music; music therapy; repertoire selection; deep learning

1. Introduction

In competitive sports, an important way to improve the level of athletes is to conduct scientific and effective training. Athletes spend most of their careers training to improve their physical fitness, playing skills and confidence. Footballer Cristiano Ronaldo and basketball player LeBron James have maintained high competitive levels for nearly two decades of their career, thanks to their hard training. In the training process of athletes, we should step by step, balance the training load and recovery, reasonably match the load amount and load intensity, and arrange the training content according to the training purpose and the characteristics of sports [1]. The degree of training load should not be blindly increased. Among these things, training load, is definitely an art than a science. It's the total quantity of activity that perform, often calculated over the course of a week, and it may be calculated using two metrics: frequency and effort. Measurement of duration is simple. Intensity multiplied by period is known as training volume. The exercise concentration rises when people do out harder and/or faster. If assessed throughout period, longer working can also represent the cumulative load. The exercise intensity for the week is calculated by summing the normal training loads across seven days. Large amount of exercise and high intensity training can make athletes maintain a high level of competition, but excessive training is easy to make athletes 'recovery period longer, or even accidental injury, causing irreparable consequences to athletes' career [2]. Training load taken into account a number of variables that have an impact on the exercise intensity as well as faster recovery, including pulse rate while exercise, workout length, and personal characteristics like race, gender, size, and obesity.

It is particularly important to detect the athletes and evaluate the physical function of the professional athletes [3]. Through the analysis of the physiological and biochemical indicators of the athletes, we can understand the ability of the training load, the current physical function state and the scientific and effective training arrangement of the athletes [4]. Exercise has been shown to reduces blood pressure also has favorable effects on blood fat concentrations, glucose levels balancing, as well as systemic aggravation in addition to aiding in weight management. Our fitness and outlook are both improved by all of these adjustments. On the other hand, the judgment of athletes' training load bearing status should be a comprehensive evaluation of multiple indicators and multiple levels, among which, the physical state of athletes is a very important factor [5]. According to the age, physical quality, physiological and biochemical index, and training level of the test subjects.

Cycling has risen soon since its birth. The world's first cycling race was held in France in 1968, and so far there have been over hundreds of cycling races [6]. After the founding of the People's Republic of China, cycling was officially listed as a sport at the first National Army Games in 1952, and there have been national cycling competitions every year since 1957 [7]. In recent years, China's cycling event has made new achievements in international competitions, winning its first gold medal at the Rio Olympic Games. The national cycling team has accumulated a lot of data in the years of training and competition, and how to use these data to help athletes improve their training effect and achieve new Olympic success is a problem to be solved [8].

With the continuous development of data mining technology, the informatization of competitive sports has become a trend. It has become normal to assist athletes to train, help coaches to make decisions and improve the competitive level of sports circles. As an interdisciplinary application technology, data mining covers the fields of statistics, machine learning, fuzzy mathematics and other scientific fields. Data analysis and knowledge extraction according to specific problems. Its application in the field of sports can significantly improve the level of sports competition and management level [9–11].

Apply data mining technology to cycling projects, the athletes of physiological and biochemical indicators, competitive level indicators and the relationship between training load ability, build a cyclist training load prediction model can not only help coaches and athletes to evaluate training schedule, but also can play the role of auxiliary analysis, decision support.

This paper screened and analyzed the data, and according to the suggestion of the national cycling team coach, 25 key factors affecting the maximum oxygen intake and the training load capacity were selected according to determine the functional threshold power and blood oxygen saturation. In this paper, the BP neural network is used to construct the prediction model. Then, this paper improves the selection operator of the adaptive genetic algorithm and adds the optimal conservation strategy based on the use of ranked selection, which ensures the population diversity while improving the efficiency of genetic operation. The initial weights and thresholds of BP neural networks are optimized using modified adaptive genetic algorithms to easily trap the network into local minima and improve the convergence rate. Through experimental validation, the cyclist training load prediction model constructed using the optimized BP neural network has good accuracy. Finally, the experimental environment is introduced, and the experimental results are analyzed.

2. Related work

Load prediction has emerged as a critical task in various domains, including building energy management, athletic training, unmanned aerial vehicles (UAVs), cloud computing, and smart grids. Recent advances in machine learning (ML) have enabled more accurate and efficient prediction models, thereby improving operational decisions and resource allocations.

In building energy management, researchers have developed ML-based pipelines to achieve high-precision baseline load forecasts. For instance, Campodonico Avendano et al. [12] introduced pipelines employing sliding-window training schemes for hour-ahead baseline load prediction, identifying optimal training window sizes and algorithms such as Extra Trees. By doing so, they significantly improved the estimation of offered flexibility for different building categories, thus enabling fair compensation and effective demand response strategies. In sports science and physical education, predicting training load is crucial for optimizing athletes' performance and preventing overtraining. Cheng [13] proposed a BP neural network model enhanced with a genetic algorithm (BPNNGA) for physical education training load prediction, achieving an accuracy of up to 99%. In the aerospace sector, Hu et al. [14] employed a BP neural network model to predict flight loads on large UAVs, validating the model's robustness through error analysis. The application of load prediction extends to cloud computing, where accurate workload forecasting can drive efficient resource management. Zuo et al. [15] addressed the few-shot workload prediction problem using a mixed contrastive transfer learning approach. This method integrates generated samples with source domain data, improving feature representation and reducing prediction errors by over 20%. In the realm of smart grids, Xiang et al. [16] and Ziti et al. [17] focused on power load forecasting. Foulard et al. [18] proposed a CNN-based approach that leverages weather conditions and historical load data, outperforming traditional methods. Similarly, Ziti et al. [17] introduced a hybrid deep learning framework that combines CNN, LSTM, and Transformer models, achieving a load characteristic identification accuracy of 0.945 in low-voltage scenarios. In the automotive domain, Foulard et al. [18] developed a data-driven load prediction model for vehicle components, further emphasizing the versatility of ML-based forecasting methods across diverse fields. In athletic contexts, beyond the work of Cheng [13], Zheng [19] examined nonlinear regression models coupled with ultrasound data for athletic load prediction. Their comparative study found that BP neural networks outperformed other estimation methods, affirming the efficacy of intelligent modeling techniques in guiding training plans and balancing workout intensities.

Collectively, these studies underscore the effectiveness of ML-based models—particularly BP neural networks and their optimized variants—in various load prediction tasks. Improvements such as sliding-window training schemes [12], genetic algorithm optimization [13], transfer learning for few-shot scenarios [15], and hybrid architectures integrating CNN/LSTM/Transformer components [17] have all contributed to enhanced accuracy and robustness. However, direct comparisons with other optimization methods (e.g., particle swarm optimization, simulated annealing) remain limited, indicating an opportunity for future research. The insights gained from these diverse applications can guide the refinement of load prediction techniques, ensuring that chosen methodologies align with the unique demands and constraints of each domain.

3. Research on exercise training load prediction based on improved genetic algorithm

3.1. Multi-layer BP neural network and its algorithms

3.1.1. Multilayer BP neural network

In the mid-1980s, a scientific group led by David Rnmelhart was a multi-layer forward feedback network with supervised learning. Its core idea was to reduce the error of network output and expected values by adjusting the weights and thresholds of the network. Structurally, it is divided into input layer, output layer and hidden layer. The hidden layer of layer BP neural network can be more than one layer, and the input and output layer is one layer. Its topology is shown in **Figure 1**. The number of neurons in the input layer is determined according to the dimension of the input signal. The problem to study determines the number of neurons in the output layer, while the number of neurons in the hidden layer are determined according to the complexity of the specific problem and the number of neurons in the input and output layer.

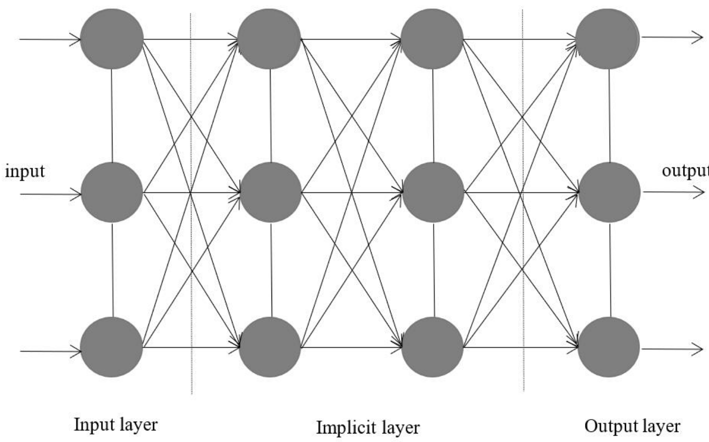


Figure 1. Multilayer BP neural network.

The process of BP neural network training is divided into the forward propagation process of input signal calculation and summing through neurons in each layer and the reverse feedback process of output error function. Forward propagation is the process of storing and processing input data before it is pushed forwards across the system to produce a result. In a deep network, the hidden units receive the information from the incoming level, analyze it using the input vector, and then send the results to the output units or subsequent levels. This article's goal is to bring users' awareness to reverse feedback, also known as Bottom-up feedback, that travels through workers to their supervisors and is a potent aspect of responses that has largely gone unnoticed as well as neglected. Through the alternation of these two processes, the connection weights and threshold of the BP neural network are constantly optimized. These two processes are described as follows:

1) The input signal is propagating forward.

The topology, weights, threshold values and other parameters of the BP neural network should be determined and initialized before training the network. The forward propagation of the input signal refers to the process in which the neural network results are obtained after the input signal is processed by each layer of neurons. In this process, the input value of each layer serves as the input value of the next layer after a weighted sum and transition through the activation function, and the connection weights and threshold of the network remain unchanged.

2) Reverse signal propagation of error.

If the errors in the network output and the actual results are large, the error backpropagation begins. Error signal backpropagation is the process in which the error between the actual and desired output of the network travels layer by layer forward from the output layer. The goal of the backpropagation method is to create a multilayered convolutional multiple machine learning technique so that the network may be educated to intuitively recognize the mappings. In fields like machine learning, stochastic gradient descent techniques are frequently employed to build feedforward neural systems. The gradients of the loss effect are related towards the connection weights is easily computed. The neural network development process uses a technique called the error backpropagation optimization technique. Calculating the slope of the loss product is the primary objective. It is frequently referred as error function or else cost function. In this process, the weights of each layer of the neural network are adjusted according to the gradient value of the error, and the purpose of adjusting the weights by error backpropagation is to reduce the error and make the output of the neural network gradually close to the desired output.

3.1.2. Derivation of the BP algorithm

The BP algorithm step is mainly to determine the topology of the network first, initializing the weight and threshold of each layer of the BP neural network. Training data of the neural network is then provided as an input signal. The output of each layer is calculated, and the error between the actual output of the neural network and the expected value is calculated through the error function. The expected value is determined in statistical as well as likelihood assessment through calculating every conceivable result by the possibility that each event will happen, subsequently adding up all of those quantities. Owners can select the option that is most probable to provide the intended effect through evaluating anticipated prices. The algorithm ends when the error accuracy reaches the requirement, otherwise the next step is continued. The weights and thresholds of the neurons in each layer were adjusted by descent gradient method, returning to the third step to continue performing the algorithm. Utilizing distinct features identified as cost factors, this fault is assessed. A strong optimizer for minimizing estimation methods is gradients descend. It makes use of numerical algorithms to incrementally adjust values in attempt to steadily lower modeling expense and get the system closer to a correct solution. Gradient descent has the drawback that the value updating at a given time (t) is solely determined both the training data as well as gradients at that time. The previous methods done when navigating the cost space are not taken into consideration.

The main idea of the BP algorithm is to constantly adjust the weights and thresholds through the error forward feedback to reduce the error of the actual and expected output. Here, a three-layer BP neural network with the activation function as Sigmoid is used as an example to derive normal propagation and back-propagation processes. A sigmoid activation functional must be employed as well as the autoencoder will have single component if it contains necessarily related incompatible categories (modal categorization). Explaining that a network handles complicated tasks depends on the sigmoid function. This utility was used as a springboard for the identification of further functionalities that result in effective and workable approaches to controlled training in deep learning systems. Because the transfer function is distinguishable, the feedforward neural rule may describe the degree of variation in the output level in relation to a modification in a certain value (So if the load is located a number of layers below the outcome). If the training sample X has n, and the corresponding expected output is t, and the actual output result is Y, let the hidden layer have S neurons. As a consequence, a significant shift in the intake of the nonlinear function will only have a minor effect on the outcome. Consequently, the futures instruments. The Vanishing Gradients Phenomenon needs to constantly whenever gradient-based methods are used to teach machine learning.

First, the positive propagation of the input signal:

The output of the i-th neuron in the hidden layer transformed by the activation function is expressed as the formula:

$$net_i = \sum_{g=1}^{n} w_{ig} x_g - \theta_i \tag{1}$$

$$a_i = f\left(\sum_{i=1}^{p} x_j w_{ig} - \theta_g\right) \tag{2}$$

Put Equation (1) into Equation (2) for $a_i = f(net_i)$, f is the Sigmoid function.

The output of the g-th neuron in the output layer is:

$$y_g = f\left(\sum_{r=1}^{n} a_r w'_{gr} - \theta'_r\right) \tag{3}$$

To make $net_i = \sum_{r=1}^{n} a_r w'_{gr} - \theta'_r$, then Equation (3) is converted to Equation (4):

Then, the error function is the Equation (4):

$$E(w, \theta) = \frac{1}{2}\sum_{g=1}^{n} \left(t_g - y_g\right)^2 \tag{4}$$

3.2. Optimization of the BP neural network by improved adaptive genetic algorithms

3.2.1. Adaptive genetic algorithm

Genetic manipulation refers to the method of creating the next generation population, including selection manipulation, crossover manipulation, and variation manipulation. The selection operation is to select the higher fitness individuals from the population according to the specific rules to inherit them to the next generation. Individuals with high fitness have better genes and can retain excellent characteristics based on their reproductive offspring. Common selection operators include roulette methods, local competition mechanism, and ranking selection. The greatest well and often employed roulette method digitally is the Martingale technique. The method's basic premise is that to raise the stakes after each defeat in order to recover the losses whenever they finally win. And then resume gambling using the original stake. Disruption, manipulation, and seeming competitiveness are the three main processes of competitors. Selecting the finest among a number of steps (that display unpredictability) based on certain evaluation metrics is described as ranking and selecting. It is necessary to replicate the steps in order to learn further about people because they are unpredictable and typically very complicated. Cross operation is an important process of genetic fabrication, by exchanging the genes of two chromosomes to create new chromosomes, the position of two chromosomes to exchange genes through the crossover probability decide. As organisms reproduce, new individuals may combine all the advantages of the parent, and crossover can create higher fitness chromosomes. Chromosome crossing method mainly consists of single-point crossing and multipoint crossing. On the parental organisms' strand, a crossing point is chosen. The two primary entities exchange all information in the creature chain after that moment. A variation on the one-point crossovers described as multi point crossing involves swapping alternate portions to develop novel descendants. The concepts of genomic networks form the foundation of evolutionary algorithm (GAs), which are simulated annealing techniques. They hunt for the best way to address the assessment (fitness) parameter of an optimization approach. GAs only makes utilize the efficiency scaling factor while dealing with numerous alternatives concurrently. According to the analysis of the principle of the genetic algorithm, the flow chart of the algorithm is shown in **Figure 2**. The advantages of genetic algorithm are probabilistic by definition, a bigger collection of potential solutions, needs little knowledge, images of DNA made from chromosomes, parallelism, offers a variety of ideal options, and so on. The specific steps are as follows:

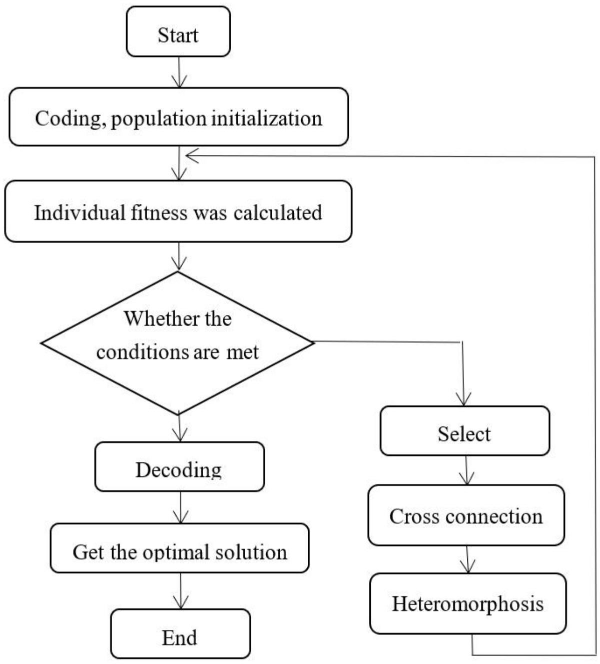


Figure 2. Genetic algorithm process.

The Adaptive Genetic Algorithm (AGA) is an improved genetic algorithm proposed by Srinvivas et al., where the cross probability and variation probability can be automatically changed according to the fitness of the chromosome. In the current study, an adaptive genetic algorithm (AGA) with a genetic operators dubbed GB-crossover is suggested. This selection mechanism combines the standard crossovers of GA with the optimal global placement of PSO in a novel way. Individuals with high quality genes reduce the probability of crossing and variation, while individuals with low fitness cross and variation with higher probability, so as to increase population diversity while retaining high quality chromosomes. The adaptive genetic algorithm also solves the problem of early puberty to some extent while ensuring the population diversity.

3.2.2. Optimization of the BP neural network based on improved genetic algorithms

This paper optimizes the BP neural network using an improved adaptive genetic algorithm from the perspective of initial weights and threshold. At present, the initial weights and thresholds are usually generated in a random way, which lacks basis, and the network performance cannot be guaranteed. Once the value is improper, the network may fail to converge or easily fall into local minima. This section uses the specific approach to optimize BP neural networks using modified adaptive genetic algorithms as follows, where the selection operation of the adaptive genetic algorithm in a general choice roulette way, while this paper proposes a simple and effective selection operator.

(1) Coding and the generation of the initial populations.

Since the initial weights and thresholds of BP neural network are numbers within the range of (0, 1), using the binary coding method will make the chromosome very long, and the efficiency of the algorithm will be greatly reduced, this paper chooses to use real coding. Real-number encoding not only eliminates the decoding work, but also facilitates the calculation of genetic operations. In this paper, the number of weights and thresholds of the networks is up to 293, and here we introduce the specific coding process with a neuronal network with a topology of 3-2-1. In this neuronal network, the connection weights between input layer and hidden layer neurons are separately: w_{11}^{1},w_{12}^{1},w_{21}^{1},w_{22}^{1},w_{31}^{1},w_{32}^{1}, The connection weights between hidden layer neurons and output layer neurons are w_{11}^{1}, w_{21}^{1} respectively, and the threshold of hidden layer and output layer neurons is θ_1,θ_2,θ_3, where: w_{11}^{1},w_{12}^{1},w_{21}^{1},w_{22}^{1},w_{31}^{1},w_{32}^{1}, w_{11}^{1},w_{21}^{1} and θ_1,θ_2,θ_3 are all numbers between [0,1], and these weights and thresholds are arranged in order to complete the real encoding. The N chromosomes (also known as individuals) are randomly generated by this encoding method, and the N chromosomes form the initial population.

The fitness function should be inversely proportional to the sum of squares of the network output error. The smaller the error, the greater the fitness. The calculation formula of individual fitness is:

$$f(x) = \frac{1}{E(x)+1} \tag{5}$$

(2) Improve the selection operator.

In the process of population selection, because the roulette method randomly selects individuals according to the proportion of individual fitness to the total population fitness, individuals with higher fitness may be missed. The main purpose of the optimal preservation strategy is to select the optimal individual in the population, but only consider the optimal individual and abandon the general fitness individuals, resulting in a single population structure and prone to local convergence. In this paper, the optimal saving policy is introduced based on the sorting and selection method, and the operation steps of improving the selection operator are as follows:

First, initialize a population, rank the individuals in the order from fitness to large; then divide the arranged individuals into four segments, increase the quality of the four segments, select the proportion of 0.4, 0.6, 0.6, 0.8 and 1, and select the individuals of the previous step into the new population, and the number of individuals in the new population remains unchanged. Use a graphical (**Figure 3**) representation for the above steps, each grid represents a chromosome, and the numbers in the grid represent the fitness values:

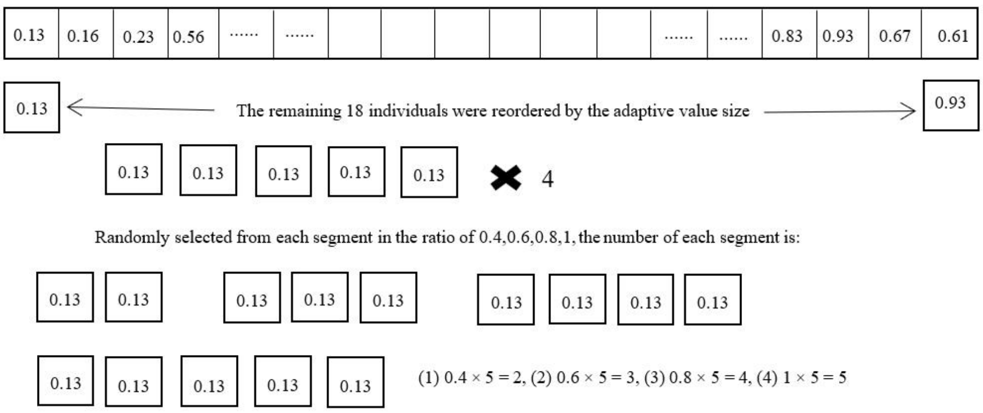


Figure 3. Improved algorithm for sub operator initial population weight selection.

The number of individuals is temporarily selected through the process above. As the number of individuals lost by the selection operation is 6, one individual is randomly selected at a time in the fourth paragraph, and a total of 6 individuals are selected. The selected individual was generated by inserting the tail of the selected individual in step 5 to obtain a new population. The improved selection operator selects other individuals while preserving the optimal individuals to ensure the population diversity. The operation is simple to calculate the fitness and then use some basic operations such as sorting, grouping and inserting. The roulette method needs to calculate the fitness of the individual, then the size of the proportion of the individual fitness and find the selection probability, so the calculation is larger. In conclusion, improving the proposed selection operator can theoretically improve the selection ability, and the convergence will also be improved because of the computational reduction.

3.3. Selection of predictor indicators of exercise training load based on an improved genetic algorithm

3.3.1. Index selection

The training load prediction model of cyclists needs to take the physiological and biochemical data of athletes, competitive level indicators and training content as the basis to predict the load degree of athletes' training, and to judge whether the currently arranged training amount is reasonable. By foreseeing upcoming use of goods that the company will transport or supply, demand forecasts reduce operational danger. Relative price, environment as well as supply reaction assessment, including prediction for solar electricity are examples of methods. When creating the load forecasting models, these variables are crucial. The historical data on load patterns, the climate, ambient temperature, wind direction, annual monthly data, financial activities, and land features are only a few of the many variables that influence prediction. Therefore, the most important purpose of this chapter is to build a training load prediction model of cyclists through the prediction algorithm. The first step of building the model is to analyze and process the data and select relevant indicators, divide the screened data into training set and test set, and use the data of the test set to verify the accuracy and performance of the trained athlete training load prediction model.

3.3.2. Sources of date

The experimental data of this paper adopts the data uploaded in the "Bicycle Team Training Analysis System", which is the data generated in the training and physiological and biochemical examination of the national cycling team over the years. The data required by the experiment are divided into basic data of athletes, training data and physiological and biochemical data (**Figure 4**). Based on the ID of the athlete and the training date of the training data, we find out the other data of the athlete within the week.

After screening, we obtained a total of 5374 complete and valid data points. To ensure transparency and credibility, we further analyzed the composition of these data. Among them, approximately 45% originated from the basic data of athletes (e.g., height, weight, age, gender), 35% were derived from training-related records (e.g., daily training volume, training intensity, average speed, and average power during sessions), and the remaining 20% were from physiological and biochemical indicators (e.g., hemoglobin, erythrocyte count, creatine kinase levels, maximal oxygen uptake, functional threshold power, and blood oxygen saturation). Each data record corresponds to a unique athlete-session instance, ensuring representativeness across different time periods and conditions. This distribution aligns well with the recommendations from the national cycling team coaches, ensuring that our sample adequately reflects the multifaceted factors influencing cyclists' training load capacity.

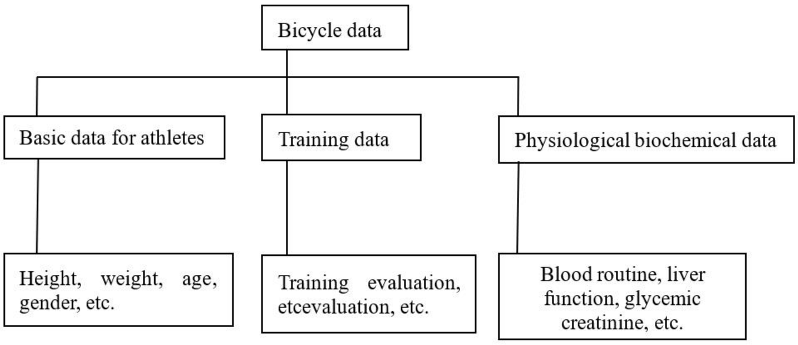


Figure 4. Cycling team data category.

3.3.3. Factors affecting the training load

In the process of predicting the training load of athletes, the factors affecting the training load should not only be reasonable but also representative. Combined with the characteristics of the experimental data, the representative factors affecting the training load in **Figure 5** are selected according to the suggestions of the national cycling team coach:

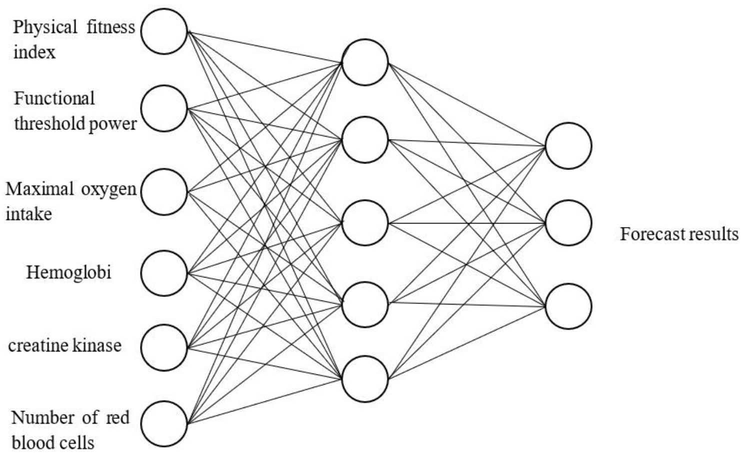


Figure 5. Network structure topology map of the training load prediction model.

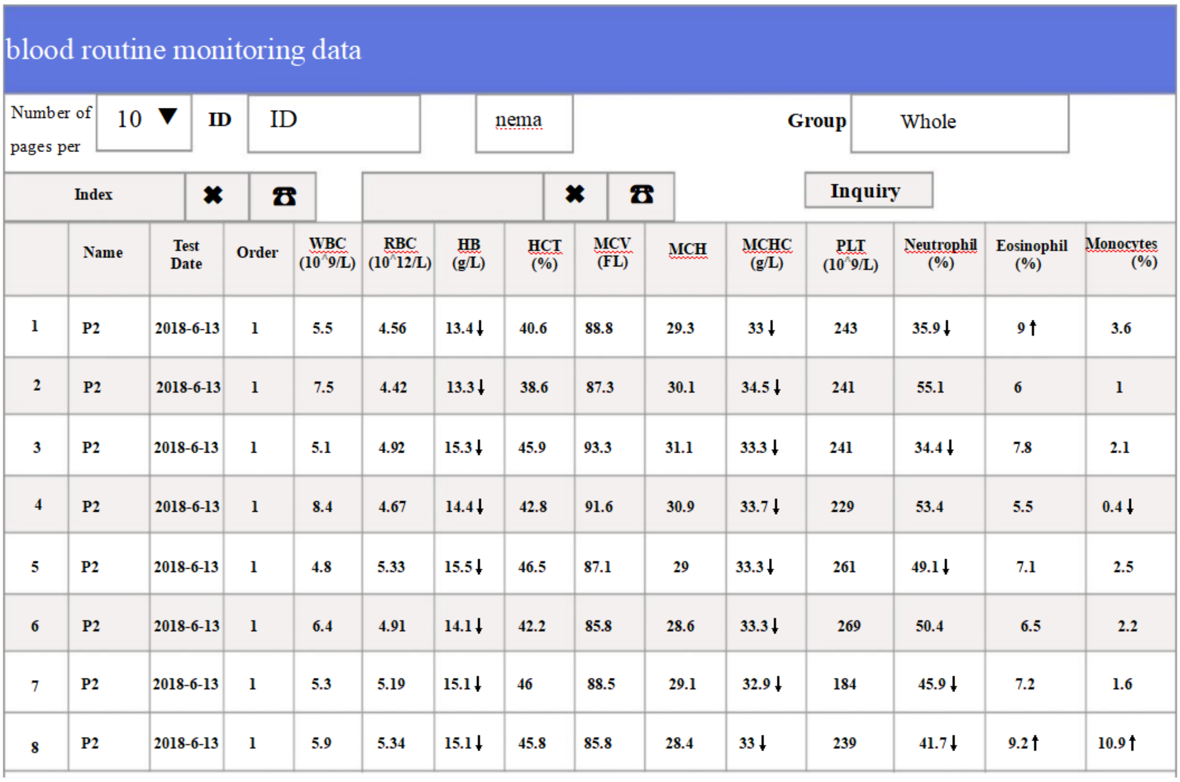
1) Height, weight, age, gender, and physical fitness index of the athletes.
2) Functional threshold power (FTP). FTP is a reliable outdoor test-based indicator for evaluating physical capacity. Nevertheless, care must be exercised whenever comparing FTP and LT because the bias among the two indicators appears to depending on the individual's level of fitness. Using the greatest previous 20-min

performance number, one may calculate FTP. For the FTP, increase the figure by 95%. Additionally, a successfully accomplish 45–60-min maximum power can be used to predict FTP. The FTP in cycling training is the maximum average power obtained in a vigorous and stable ride within 1 h.

3) Maximal oxygen intake. Maximum oxygen intake is an important indicator of the aerobic capacity of reactive athletes.

4) Physiological and biochemical indicators. There are a lot of physiological and biochemical data in the data, including blood routine, glycemic creatinine, liver function, urine routine, testosterone measurement and other major items, and each of them contains dozens of testing indicators. According to the advice of experts and referring to the research of others, hemoglobin, erythroid cell number and creatine kinase were selected in the physiological and biochemical data.

This paper takes the data of 8 athletes as an example, and the statistics of the routine blood test of athletes are as shown in **Figure 6**.



blood routine monitoring data

Number of pages per 10 ▼ ID ID nema Group Whole

Index ✖ ☎ ✖ ☎ Inquiry

	Name	Test Date	Order	WBC (10^9/L)	RBC (10^12/L)	HB (g/L)	HCT (%)	MCV (FL)	MCH	MCHC (g/L)	PLT (10^9/L)	Neutrophil (%)	Eosinophil (%)	Monocytes (%)
1	P2	2018-6-13	1	5.5	4.56	13.4↓	40.6	88.8	29.3	33↓	243	35.9↓	9↑	3.6
2	P2	2018-6-13	1	7.5	4.42	13.3↓	38.6	87.3	30.1	34.5↓	241	55.1	6	1
3	P2	2018-6-13	1	5.1	4.92	15.3↓	45.9	93.3	31.1	33.3↓	241	34.4↓	7.8	2.1
4	P2	2018-6-13	1	8.4	4.67	14.4↓	42.8	91.6	30.9	33.7↓	229	53.4	5.5	0.4↓
5	P2	2018-6-13	1	4.8	5.33	15.5↓	46.5	87.1	29	33.3↓	261	49.1↓	7.1	2.5
6	P2	2018-6-13	1	6.4	4.91	14.1↓	42.2	85.8	28.6	33.3↓	269	50.4	6.5	2.2
7	P2	2018-6-13	1	5.3	5.19	15.1↓	46	88.5	29.1	32.9↓	184	45.9↓	7.2	1.6
8	P2	2018-6-13	1	5.9	5.34	15.1↓	45.8	85.8	28.4	33↓	239	41.7↓	9.2↑	10.9↑

Figure 6. View page for blood routine test data.

These 25 factors were selected based on recommendations from national cycling team coaches, existing literature on physiological determinants of endurance performance, and preliminary correlation analyses conducted on the available dataset. Each factor was chosen for its known or hypothesized association with aerobic capacity, muscular endurance, or energy metabolism. While all selected factors contribute to model performance, their relative impact varies. Preliminary sensitivity analyses (not shown in this paper) suggest that parameters like maximum oxygen uptake (VO_2max), functional threshold power (FTP), and certain blood biomarkers (e.g., hemoglobin levels) have higher weights in determining training load predictions.

Future work will include a formal feature importance analysis to quantitatively assess each variable's contribution.

In addition to the physiological and training-related indicators, environmental conditions such as temperature, humidity, and altitude during data collection may also influence the athletes' training load capacity. For this study, the majority of training sessions were conducted in an indoor training facility maintained at approximately 20 ℃–22 ℃ and 50%–60% relative humidity, with an altitude near sea level. Although these conditions were relatively stable, slight variations still occurred over time. Such environmental fluctuations can affect thermoregulation, cardiovascular strain, and perceived exertion, potentially altering the relationship between measured indicators and actual training load. Future iterations of the model could incorporate these environmental factors as additional inputs, providing more accurate and contextually relevant predictions.

3.4. Interpretation

3.4.1. Experimental environment setting

In this paper, modified-adaptive genetic algorithm is used to optimize BP neural network to model the training load. By comparison with the traditional BP neural network model (hereinafter referred to as BP model) and the standard adaptive genetic algorithm optimization of BP neural network model (hereinafter referred to as AGABP model), the improved genetic algorithm optimization of BP neural network model (hereinafter referred to as IAGABP model) in training load prediction is analyzed. A search heuristic that is frequently employed to produce helpful answers for optimizing as well as exploration issues is the genetic algorithm (GA). It creates answers to objective function utilizing methods including transmission, variation, choosing, as well as recombination that are influenced by evolutionary biology. In order to address the issues of slow convergence density and reduced precision brought on by the initial guess loads of BP machine learning, the enhanced iterative method is utilized to optimize the values of BP neural systems.

In order to verify the prediction model, the data screened in subsection 3.2.2 is divided into two parts, with one part of the training data selected as training samples and the other part as test samples. The BP neural network is set to the network topology of 25-10-3 from the above introduction with a target accuracy of 0.001, a maximum number of 1000 and a learning rate of 0.1. Because users are generating minuscule adjustments to the parameters in the system, education will go extremely gradually if the acquisition speed is set too minimal. The loss function may exhibit unwanted divergence behavior if the training speed is set too enough. The initial population size of the genetic algorithm is set to 40, the maximum evolutionary algebra is 25, and the crossover and variation probability are determined by the adaptive formula. The prediction results are divided into three categories, namely large training, moderate training and small training.

To evaluate the robustness and stability of our model under different hyperparameter settings, we additionally conducted a sensitivity analysis. We varied the learning rate (0.05, 0.1, 0.15), population size in the genetic algorithm (20, 40, 60), and maximum evolutionary generations (15, 25, 35) to observe changes in convergence speed and prediction accuracy. Preliminary results show that a learning rate of 0.1 and a population size of 40 yield a good balance between convergence rate and accuracy, while increasing the population size beyond 60 slowed down convergence without notable accuracy improvements. Similarly, increasing the maximum evolutionary generations above 25 offered diminishing returns in accuracy gains.

3.4.2. Interpretation

To verify the effectiveness of the BP neural network improvement method, the experiment first compares the error accuracy in training and the fitness value of population optimal individuals, and then compares the accuracy of different prediction models in the validation using test samples.

The plots of error accuracy between the BP neural network model and IAGABP model are shown in **Figures 7** and **8**. The traditional BP neural network model iterates the target accuracy by 0.001 by about 1000 times, while the IAGABP model achieves the target accuracy by about 450 times. The IAGABP model can achieve the target accuracy faster than the BP network model due to the optimal initial weights and threshold by improved genetic algorithm for the BP neural network. Therefore, optimizing the BP neural network using an improved adaptive genetic algorithm has obvious advantages and stability in the convergence process.

The fitness contrast maps of the AGABP model and the IAGABP model based on the standard adaptive genetic algorithm are shown in **Figure 7**. The fitness of the AGABP model has not been stable at 16 generations, and the fitness curve of the IAGABP model stabilizes after 13 generations. By improving the selection operator, the optimal chromosome fitness in the IAGABP model is 0.18 higher than the AGABP model, which shows that the population optimization ability of adaptive inheritance is improved and the algorithm can search for excellent weight threshold.

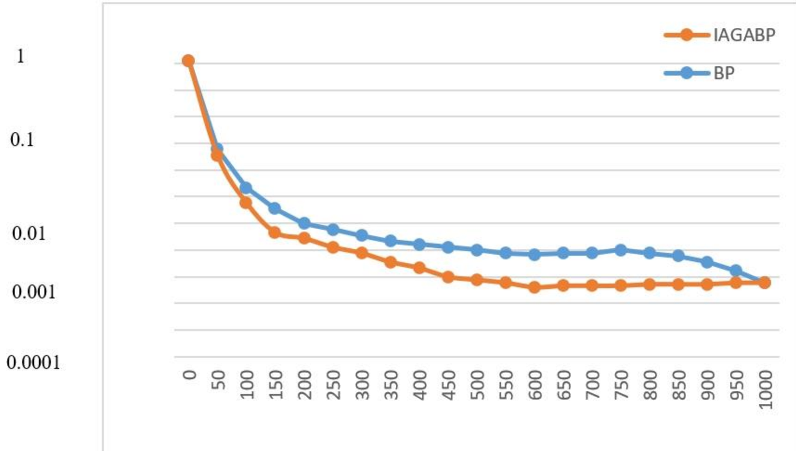


Figure 7. Error trend chart.

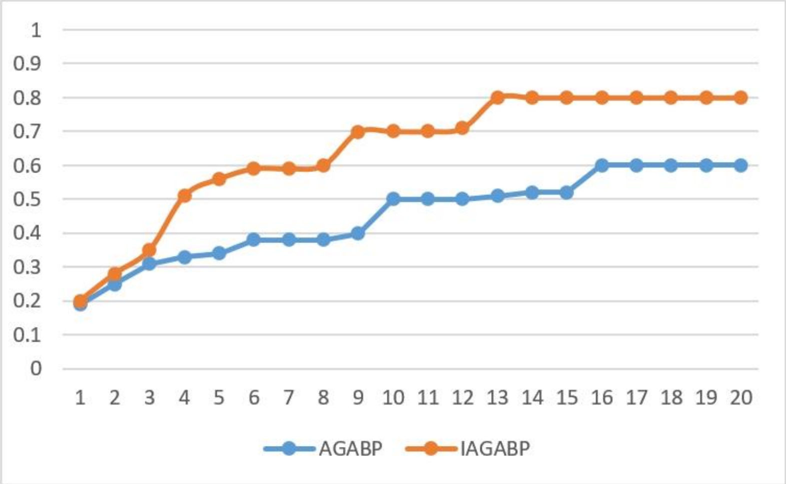


Figure 8. Adaptation trend chart.

According to the test results of the three models of the test samples are shown in **Table 1**, the first type is the accuracy with small training volume, the second type is the accuracy with moderate training volume, and the third type is the accuracy with large training volume. It can be seen from the results that the prediction model optimized by the adaptive genetic algorithm was improved by 12.73% over the BP neural network model, while the IAGABP model improved the selection operator by 1.91% better than the AGABP model. From the analysis of the above experimental results, the paper shows that the BP neural network prediction model optimized by improving the genetic algorithm has better performance and relatively high accuracy, and meets the requirements of training load prediction for athletes.

Table 1. Prediction accuracy of the different models.

Training of the load prediction model	Forecast accuracy	First kind	Second kind	Third class
BP	78.64%	80.31%	81.74%	73.16%
AGABP	91.37%	90.63%	92.48%	91.61%
IAGABP	93.28%	90.57%	95.37%	91.67%

To further validate the advantages of the improved genetic algorithm (IAGABP), we conducted additional experiments comparing our method with other common optimization algorithms used for neural network weight optimization, such as Particle Swarm Optimization (PSO) and Simulated Annealing (SA). We integrated both PSO and SA separately into the BP neural network framework, forming two comparative models: PSOBP and SABP. Using the same dataset and evaluation metrics, PSOBP and SABP achieved prediction accuracies of approximately 91.02% and 90.47% respectively, which are lower than the 93.28% accuracy achieved by the IAGABP model. In terms of convergence speed and stability, IAGABP also demonstrated more rapid attainment of the target accuracy and exhibited fewer oscillations during the training process. These results confirm that the improved selection operator and adaptive strategy in our IAGABP approach provide competitive advantages over other widely employed optimization methods, thus validating the effectiveness and robustness of the proposed method. The quality of the two athletes is analyzed, and the difference between the physical quality of the two athletes can be intuitively seen through the radar map. As shown in **Figure 9**, Xu Chao's maximum oxygen intake is higher than Zhang Lei, indicating that he has better aerobic exercise ability.

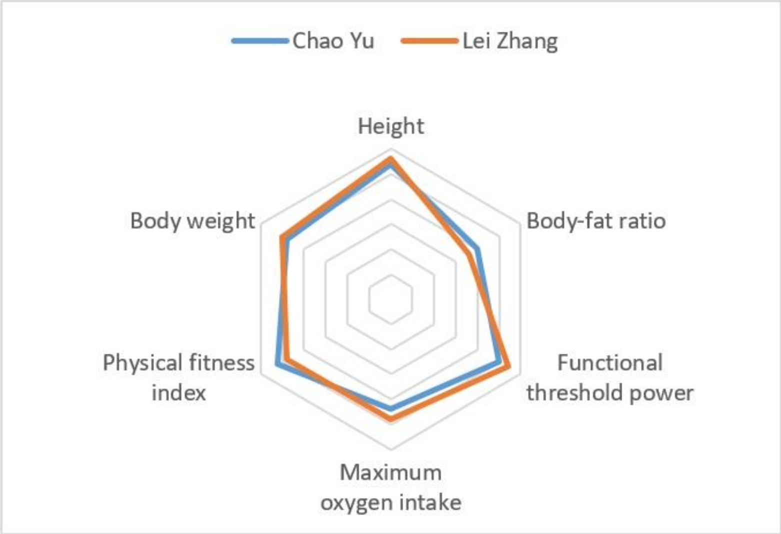


Figure 9. Physical fitness radar map of athletes.

The bar chart directly shows the difference between the performance of the two athletes in the same training program. As shown in **Figure 10**, Zhang Lei's average speed and average power are higher than that of the other two people in the training program.

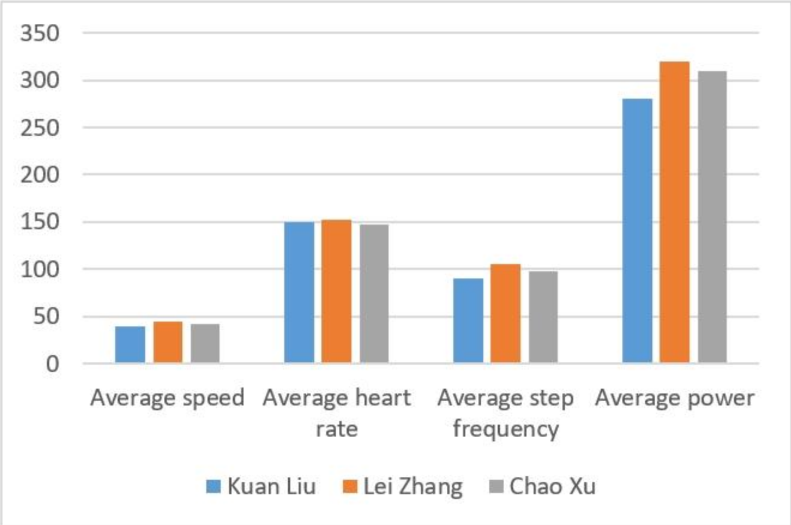


Figure 10. Bar chart of athlete training comparison.

Through the line chart, we can see the change pattern of physiological and biochemical data of athletes over a period of time. The changes of blood hemoglobin, platelets and neutrophils in blood routine tests in June 2021 are shown in **Figure 11**.

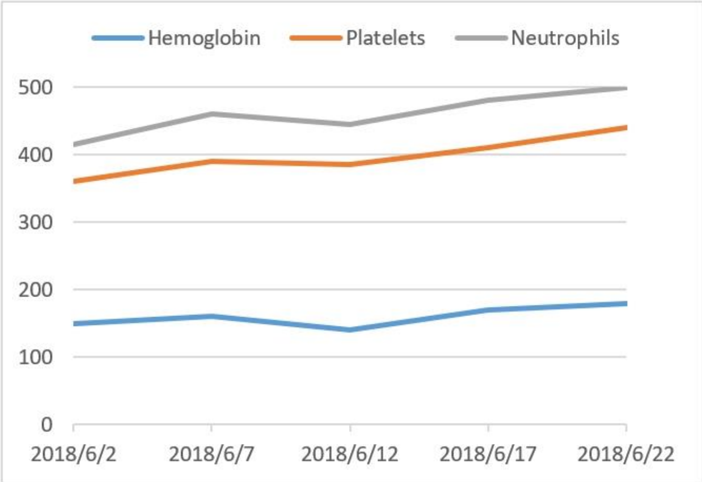


Figure 11. Line chart of physiological and biochemical changes in athletes.

4. Conclusion

After years of development, China has made great progress, but with the development of big data, many sports have applied data mining technology, which brings new challenges to the further good results of cycling projects. The national cycling team has accumulated years of training and competition data. It is important to find out the relationship between the athletes' training load ability and physical function status to improve the training effect of the cycling team. With the data of the national cycling team, we propose a prediction model of cycling training load. First, the data of the cycling team was screened, and 25 factors affecting the training load ability of athletes were selected from the data. Due to a nonlinear relationship between many factors and the training load ability, the model should also be applied to the system. This paper selects BP neural network as the basic algorithm of the prediction model. In order to improve the accuracy and convergence rate of the model, we use the adaptive genetic algorithm to optimize the initial weight and threshold of the BP neural network, and improve the selection operator of the adaptive genetic algorithm. The athlete training load prediction model constructed using a modified BP neural network optimized by the adaptive genetic algorithm achieved an accuracy of 93.28%. Although the accuracy of the athlete training load prediction model in this paper, the development of the system has also been preliminarily completed, can complete the current user requirements. However, due to the limited time and my limited ability, there are still imperfect topics, improvements are needed later on, mainly from the verification results of the athletes' training load prediction model, we can see that although the accuracy of the model has reached a good level. But there is room for optimization, the influencing factors selected in the model are the physiological, biochemical indicators and physical function indicators of the athletes. The impact of recent training by athletes and the training environment such as altitude and temperature was not considered. Therefore, in the subsequent research, we should expand the range of the influencing factors selected in the model, and take into account the recent training of the athletes and the impact of the training environment, such as altitude and temperature. At the same time, in the environment actively advocated by national fitness in China, cycling has become a popular choice for ordinary people to exercise. It should also design a bicycle training load prediction model for the situation of non-professional people to help them carry out scientific and reasonable exercise and achieve the purpose of effective fitness.

In the future, we plan to integrate environmental variables such as temperature, humidity, and altitude fluctuations into the prediction model. Advanced sensor technologies and wearable devices can continuously capture these environmental parameters alongside physiological data, enabling dynamic adjustments of training loads in real-time scenarios. Incorporating real-time feedback loops, where the model continuously updates its predictions as environmental or physiological conditions change, would allow coaches and athletes to instantly modify training intensities or volumes. Additionally, developing a mobile or wearable application interface would facilitate on-the-spot decision-making and more personalized training guidance, ultimately enhancing the model's practical value for the sports science community.

Author contributions: Conceptualization, BL and JC; methodology, QW; software, WH; validation, BL, JC and QW; formal analysis, WH; investigation, QW; resources, JC; data curation, BL; writing—original draft preparation, BL and JC; writing—review and editing, JC; visualization, QW; supervision, WH. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. Sports Medicine, 2014, 44(Supplement 2): 139–147.
2. Aleksina AO, Chemova DV, Ivanova LA, et al. The Main Directions in Informatization of the Sphere of Physical Culture and Sports Services. In: Popkova, E., Ostrovskaya, V. (eds) Perspectives on the Use of New Information and Communication Technology (ICT) in the Modern Economy. ISC 2017. Advances in Intelligent Systems and Computing, vol 726. Springer, Cham. https://doi.org/10.1007/978-3-319-90835-9_56
3. Popkova EG, Ostrovskaya VN, (eds). Perspective on the Use of New Information and Communication Technology (ICT) in the Modern Economy. Springer, Cham, 2017, pp. 473–479.
4. Ruppert D. The Elements of Statistical Leaming: Data Mining, Inference, and Prediction. Publications of the American Statistical Association, 2010, 99(466): 567–567.
5. Schumaker RP, Solieman OK, Chen H. Sports Knowledge Management and Data Mining. Annual Review of Information Science & Technology, 2010, 44(1): 115–157.
6. Sliz BA. An Investigation of Three-point Shooting through an Analysis of NBA Player Tracking Data. 2017, https://doi.org/10.48550/arXiv.1703.07030
7. Rosli CMFCM, Saringat MZ, Razali N, et al. A Comp amative study of Data Mining Techniques on Football Match Prediction. Journal of Physics: Conference Series, 2018, 1020: 1742–1753.
8. Carpita M, Sandri M, Simonetto A, et al. Discovering the Drivers of Football Match Outcomes with Data Mining. Quality Technology & Quantitative Management, 2015, 12(4): 561–577.
9. Vilela T, Portela F, Santos MF. Towards a Pervasive Intelligent System on Football Scouting- A Data Mining Study Case. World Conference on Information Systems and Technologies. Springer, Cham, 2018, 341–351.
10. Hu ZH, Zhou JX, Zhang MJ, et al. Methods for Ranking College Sports Coaches Based on Data Envelopment Analysis and PageRank. Expert Systems, 2015, 32(6): 652–673.

11. Leung CK, Joseph KW. Sports Data Mining: Predicting Results for the College Footbal1 Games. Procedia Computer Science, 2014, 35: 710–719.
12. Avendano IAC, Javan FD, Najafi B, Moazami A, Rinaldi F. (2023). Assessing the impact of employing machine learning-based baseline load prediction pipelines with sliding-window training scheme on offered flexibility estimation for different building categories. Energy & Buildings, 294.
13. Cheng X. (2022). Application of optimized bp neural network model for the training load prediction in physical education teaching. Wireless Communications & Mobile Computing, 2022.
14. Hu X, Pan X, Wang J. (2024). Horizontal tail flight load prediction method for large uavs based on machine learning. IOP Publishing Ltd.
15. Zuo Z, Huang Y, Li Z, Jiang Y, Liu C. (2025). Mixed contrastive transfer learning for few-shot workload prediction in the cloud. Computing, 107(1), 1–23.
16. Xiang S, Zhen C, Peng J, Zhang L, Pu Z. (2023). Power load prediction of smart grid based on deep learning. Procedia Computer Science, 228, 762–773.
17. Ziti C, Wei W, Wei J, Jun G, Yang L. (2024). High-precision identification and prediction of low-voltage load characteristics in smart grids based on hybrid deep learning framework. International Journal of Low-Carbon Technologies.
18. Foulard S, Fietzek R, Esbel O. (2022). Method for determining a load prediction for a component of a vehicle. US2022178789A1.
19. Zheng Z. (2022). Load prediction model of athletes’ physical training competition based on nonlinear algorithm combined with ultrasound. Contrast Media & Molecular Imaging.