www.cad-journal.net Computer-Aided Design and Applications

Exploration of Sports Data Analysis and Fitness Effect Optimization Strategies using XGBoost

Qianqi Zhao[1], Zhenzhen Bie[2] and Yang Li[3]

[1,3] Department of Sports Art, Hebei Institute of Physical Education, Shijiazhuang 050041, China, [1]2004011@hepec.edu.cn, [3]1993009@hepec.edu.cn
[2]Department of Physical Education, Tongji University, Shanghai 200092, China, 01095@tongji.edu.cn

Corresponding author: Qianqi Zhao, 2004011@hepec.edu.cn

Abstract. Compared with traditional methods, computer-aided technology can improve the efficiency, real-time performance, and accuracy of sports data analysis and present more comprehensive and multidimensional analysis results to fitness enthusiasts, helping them improve their fitness effectiveness. Therefore, this article combines machine learning algorithms to construct a model for analyzing sports data and quantitatively evaluating fitness effects. It also improves the accuracy and effectiveness of sports data analysis through monocular cameras and the k-means algorithm. At the same time, XGBoost is used to achieve quantitative evaluation of exercise effects. The experimental results show that the model proposed in this article can effectively improve the accuracy of evaluating the quality of fitness posture, and the accuracy of different exercise postures can reach over 97%. At the same time, the model has shown good performance in quantifying motion effects in different datasets and can adapt to different application environments. The application results show that the model can analyze the correlation between corresponding important features of exercise based on the exercise data of fitness enthusiasts and help fitness enthusiasts horizontally understand the gap between their own exercise status and the exercise status of others. Based on the analysis of exercise data, fitness enthusiasts can effectively adjust their exercise posture and plan to improve exercise effectiveness.

Keywords: Computer-Aided; Machine Learning Algorithms; Sports Data Analysis; Xgboost; Strategy Optimization
DOI: https://doi.org/10.14733/cadaps.2025.S4.153-166

1 INTRODUCTION

According to relevant data, the global fitness and entertainment sports centre market is expected to reach 770.692 billion yuan in 2023 and is projected to reach 1056.716 billion yuan by 2029, indicating that the global fitness industry is in a steady growth stage [1]. The Chinese fitness market has also shown strong growth momentum, with the overall market size continuing to expand and the penetration rate increasing year by year [2]. From 2022 to 2023, the penetration rate of fitness enthusiasts in China has increased from 26.5% to 27.6%, indicating a continuous increase in consumer interest and demand for fitness. With the intensification of population aging and the prominence of sub-health problems, areas such as healthy elderly care and health management will also bring new growth points to the fitness industry [3]. Therefore, in the future development process, the fitness industry needs to provide diversified and personalized services to more consumers. The analysis of fitness exercise data plays a crucial role in the current development of the fitness industry and is an important driving force for the development and reform of the fitness industry [4]. By analyzing the exercise data of fitness enthusiasts in depth, one can understand their physical condition, exercise habits, preferences, and goals, and tailor personalized training plans and dietary recommendations for them [5]. This precise matching can significantly improve fitness effectiveness and enhance user experience. Improving the ability to analyze sports data will become an inevitable trend in the development of the fitness industry [6].

In recent years, with the deepening of research on sports science, people have begun to realize that the strategies, rules, events, etc. in sports contain rich optimization ideas, which can inspire us to design more efficient and flexible search and optimization algorithms. Some scholars have conducted in-depth research on the specific modelling process of these motion-based computational intelligence algorithms, analyzing their uniqueness and novelty compared to existing optimization algorithms. In the process of exploring computer-aided sports data analysis and fitness effect optimization strategies, we realized that traditional search and optimization algorithms often seem inadequate when facing complex and changing sports data analysis and personalized fitness plan development [7]. Therefore, computational intelligence optimization algorithms, especially those inspired by nature and capable of simulating complex processes in nature, have brought unprecedented opportunities to this field. Especially when dealing with large-scale, nonlinear datasets and developing optimal training plans under strict constraints, the efficiency and effectiveness of these classic algorithms are limited. To further validate the effectiveness of these algorithms, we conducted comparative experiments for the first time between motion-based algorithms and popular algorithms such as genetic algorithms, particle swarm optimization, and differential evolution. Not only does it demonstrate strong exploration capabilities, but it also surpasses traditional optimization methods in multiple application scenarios, providing new perspectives and tools for sports data analysis and fitness effect optimization. It was found that these motion-based algorithms exhibit higher efficiency and better adaptability in processing complex and variable motion data [8]. By comparing the performance of different algorithms in processing sports data, identifying sports patterns, and developing personalized training plans. These motion-based computational intelligence algorithms, such as league championship algorithms, football league optimization, football game optimization, etc. These motion-based algorithms have shown impressive performance in unconstrained global optimization benchmark problems, especially when dealing with datasets with different features, they can quickly find high-quality solutions [9].

Sports data analysis requires fitness enthusiasts to collect a large amount of data every day. Traditional analysis methods often require a lot of manpower and resources to collect, organize, and analyze data when faced with massive amounts of data. This not only incurs high costs but also lacks effective data processing and simplification capabilities, resulting in a lengthy and inefficient data analysis process with high error rates. At the same time, traditional sports data analysis lacks intuitiveness, real-time performance, and predictability in presenting analysis results, which may be limited by data sources and may not fully reflect the state and fitness effects of fitness enthusiasts. There are limitations when dealing with complex and variable data, making it difficult to uncover correlations between data and provide reliable data support for improving fitness effectiveness strategies in the future [10]. Computer-assisted technology has excellent processing capabilities for massive data, which can quickly clean, integrate, and analyze data, providing timely and accurate information support for coaches and fitness enthusiasts. This real-time capability helps to adjust training plans in a timely manner, ensuring maximum training effectiveness. In subsequent fitness training, computer-aided systems can tailor personalized training plans for users based on their physical condition, fitness goals, preferences, and other information to improve fitness effectiveness. Therefore, this article combines computer-aided technology to construct a sports data analysis and fitness effect optimization strategy model. The main innovations of this article are as follows:

Firstly, sports data analysis requires the collection of relevant data with high precision from fitness enthusiasts. Therefore, to provide data collection capability and accuracy, this article combines monocular cameras to improve the accuracy and quality of motion detection for fitness enthusiasts.

Secondly, this article combines the k-means algorithm in machine learning algorithms to classify raw motion data, remove abnormal data from the data, achieve the goal of motion style classification, and improve the accuracy of motion data analysis.

Finally, this article quantitatively analyzes the fitness effects of fitness enthusiasts using the XGBoost algorithm and provides decision data support for improving fitness effects in the future.

2 RELATED WORK

Compared with traditional manual analysis methods, computer-aided analysis can automate tasks such as data preprocessing, feature extraction, and pattern recognition, greatly reducing manual intervention and error rates, and improving analysis efficiency and accuracy. Muzammal et al. [11] collected and analyzed real-time motion data, such as heart rate, speed, distance, etc. And provide real-time feedback and suggestions directly to coaches and fitness enthusiasts through visualization technology. With the development of computer technology, significant progress has been made in the analysis of sports data and optimization of fitness effects both domestically and internationally. In China, Ryselis et al. [12] have developed and applied an "intelligent sports training system". This system integrates advanced technologies such as big data, cloud computing, and the Internet of Things, creating a comprehensive and personalized training management platform for athletes. Real-time data of athletes during training, including heart rate, speed, distance, strength output, muscle fatigue level, etc., are collected through wearable devices such as heart rate monitoring bands, GPS sports watches, electromyography sensors, etc. Then upload these data to the cloud server for storage and analysis. Salloum and Tekli [13] also use data mining and machine learning techniques to deeply process these massive amounts of data, identifying athletes' physical changes, skill mastery, and psychological fluctuations at different training stages. For example, by analyzing heart rate variability (HRV) data, the recovery status and stress level of athletes can be evaluated, thereby adjusting training intensity and rest time to avoid overtraining or insufficient recovery.

Santos et al. [14] used video analysis techniques to finely evaluate athletes' technical movements. Provide scientific training guidance and suggestions for coaches, such as swimming efficiency and basketball shooting posture. In foreign countries, the application of sports data analysis focuses more on using advanced algorithms and complex models. Taking football as an example, artificial intelligence technology is not limited to simple data statistics such as shooting frequency and ball possession rate but rather deeply analyzes game tactics. Through deep learning algorithms, artificial intelligence can recognize players' complex behaviours such as running patterns, passing routes, and defensive positions, and predict game trends. These analysis results provide unprecedented tactical insights for Singh and Vishwakarma, helping them develop more precise and effective game strategies. In personal projects such as golf, artificial intelligence can analyze players' swing movements, ball speed, trajectory and other data, combined with physical simulation technology, simulate the changes in ball diameter under different hitting methods, and help players find the best hitting strategy. In the field of long-distance running, artificial intelligence monitors physiological indicators such as heart rate, step frequency, and stride length of athletes, and combines environmental factors such as altitude, temperature, and humidity to tailor training plans for athletes, maximizing training effectiveness and reducing injury risks. At present, the application of different technological means in sports data analysis and fitness effect optimization at home and abroad not only improves the training efficiency and competitive level of athletes but also provides more scientific and personalized fitness guidance for fitness enthusiasts.

In today's society, establishing a healthy lifestyle has become an indispensable part of improving the quality of life. This is not only related to the balance and nutrition of daily diet but also closely related to the reasonable planning of exercise to balance calorie intake and expenditure. Traditional nutrition and health consultations often rely on the professional knowledge and experience of human experts, which is expensive, time-consuming, and difficult to promote to a wider population. With the advancement of technology, people are increasingly inclined to use computer-aided means to optimize this process, especially through sports data analysis and fitness effect evaluation, to achieve more personalized and efficient health improvement. Singh and Vishwakarma [15] delved into the application of computer-aided optimization in promoting a healthy lifestyle. Despite the emergence of various electronic nutrition solutions in the market, most of them overlook the crucial step of health assessment when providing dietary plans, which is crucial for developing scientifically sound dietary and exercise plans. Especially through the development of a personalized intelligent nutrition recommendation system called PIN, the gap in automatic nutrition and health assessment, as well as progress tracking, has been filled. It not only provides precise guidance for weight management, calorie intake, and exercise recommendations but also enables continuous assessment and timely adjustment of individual health status, ensuring dynamic optimization of nutrition and exercise plans. The Progress Evaluation and Suggestion Adjustment (PERA) module is responsible for continuously monitoring the improvement of user health status and adjusting the suggestion plan based on data analysis results to ensure the effectiveness and adaptability of the plan.

3 CONSTRUCTION OF FITNESS EFFECT OPTIMIZATION STRATEGY

At present, many fitness enthusiasts complete corresponding fitness projects through APP software or record the fitness process through mobile devices, completing operations such as monitoring, recording, and analyzing fitness movements. Taking into account the actual conditions and cost of fitness equipment, this article uses a monocular camera to detect key points in the human body, capture the movement trajectory and posture of fitness enthusiasts, and then model and predict the movement data through machine learning algorithms to provide personalized training suggestions for fitness enthusiasts and improve fitness effectiveness.

3.1 Motion Data Analysis Module

The collection of exercise data for fitness enthusiasts is based on mobile devices, and the overall architecture of this module is divided into two core components: front-end data collection and back-end algorithm operation. The front end utilizes the cameras on fitness enthusiasts' mobile devices to collect real-time fitness action video data. At the same time, the results will be instantly fed back to the user in the form of voice prompts based on the data returned by the backend, and the evaluation details will be displayed through the user interface, supporting interactive operations such as page jumping, to provide a more intuitive and convenient user experience. After receiving the real-time data stream from the front end, the backend uses high-performance algorithms to perform a series of complex processing on the video data, including but not limited to action recognition, posture analysis, and comparison and evaluation with standard movements, in order to evaluate the quality of the user's fitness movements accurately. Finally, the processing results will be quickly returned to the front end, supporting subsequent display and interaction by the front end. The entire system achieves a closed-loop process from video data collection and processing to result in feedback through close collaboration between the front-end and back-end, providing users with real-time and accurate fitness movement guidance and feedback, thereby enhancing their fitness experience and effectiveness. The schematic diagram of the motion data analysis module framework is shown in Figure 1.

Due to the limitations of the mobile device camera industry, this article combines 2D and VideoPose3D models to obtain key point information about fitness enthusiasts' bodies and constructs an action classification skeleton database based on the key point information. Pre-input standardized movement postures of relevant fitness projects into the database and output a joint angle dataset composed of key points of the human body.

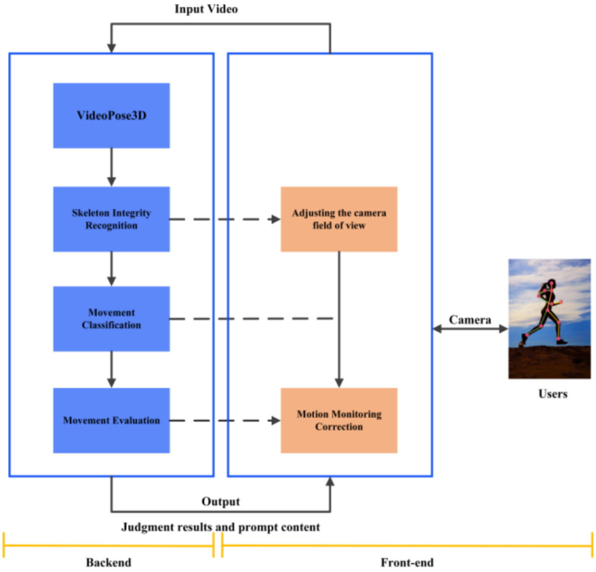

Figure 1: Schematic diagram of motion data analysis module framework.

The real-time action key points collected afterward will be compared with the standard posture key points, and the action with the highest similarity will be obtained through calculation. The calculation formula is shown in (1):

$$D = \max\left\{\sum_{n=0}^{N}[\cos(d_n - d_n^{'}] + 1\right\} \tag{1}$$

Among them, the number of actions is represented as N, the joint angles of action key points with sequence number n are denoted as d_n, and the corresponding standard action keypoint joint angles are denoted as $d_n^{'}$.

The calculation of the key point angle requires first converting its spatial coordinates into spatial vectors and then calculating the vector angle, as shown in formula (2):

$$D = \arccos[(a \times b) / (|a| \times |b|)] \tag{2}$$

Among them, the spatial vectors composed of spatial coordinates are represented as a, b.

There are differences in the angles of key points for different exercise postures, so it is necessary to calculate the direction of key points for each movement. Considering the limitation of space, this article mainly evaluates the quality of movements using common fitness exercise tablet supports. The key points for evaluating the quality of flat support movements are whether the hip and knee key points are standard. The calculation formulas for the two are shown in (3) and (4):

$$D = \left[\pi - \arccos\left(\frac{e_b \times e_y}{|e_b| \times |e_y|}\right)\right] - \arccos\left(\frac{e_l \times e_y}{|e_l| \times |e_y|}\right) \tag{3}$$

$$D = \arccos\left(\frac{g_{thigh} \times h_{shank}}{|g_{thigh}| \times |h_{shank}|}\right) \tag{4}$$

Among them, the body variable is represented as e_b, the Y-axis variable is represented as e_y, the overall leg variable is represented as e_l, the thigh variable is represented as g_{thigh}, and the calf variable is represented as h_{shank}.

Due to the influence of living environment, working environment, and personal physical conditions, different fitness enthusiasts exhibit different exercise abilities and states, as well as different exercise styles. If only one sports-style model is used for data analysis, the error rate of the results will increase. Therefore, after data collection or abnormal data processing, it is necessary to classify exercise style data and further analyze the relevant data of fitness enthusiasts based on different exercise styles. This article uses the k-means algorithm to construct data classification for different sports styles. It is a widely used clustering algorithm aimed at dividing points in the dataset into K clusters so that each point belongs to the cluster represented by the nearest mean. This algorithm iteratively finds the optimal cluster partition, making the points within the cluster as similar as possible and the points in different clusters as different as possible. Meanwhile, considering the periodicity and frequency of motion data, the Fourier algorithm is combined to extract the corresponding frequency components of the data and then classify the data.

If the exercise data of the fitness enthusiast is represented as $x=(x_0,x_1,...,x_{N-1})$, and this vector represents the corresponding number of exercises performed by the fitness enthusiast over a period of time then each data vector can be transformed by Fourier transform, as shown in formula (5):

$$X(u)=\frac{1}{N}\sum_{i=0}^{N-1}x(i)e^{-j2\pi\frac{ui}{N}} \tag{5}$$

Among them, $u=0,1,...,N-1$, the number of extracted points per cycle in the frequency domain is N.

The transformation matrix is shown in formula (6):

$$H=\frac{1}{N}\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & e^{-j2\pi\frac{1}{N}} & \cdots & e^{-j2\pi\frac{N-1}{N}}\\ 1 & e^{-j2\pi\frac{2}{N}} & \cdots & e^{-j2\pi\frac{2(N-1)}{N}}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & e^{-j2\pi\frac{N-1}{N}} & \cdots & e^{-j2\pi\frac{(N-1)(N-1)}{N}}\end{bmatrix} \tag{6}$$

Let the time-domain data matrix be represented as $X=x^T=\begin{bmatrix}x_0\\ x_1\\ x_2\\ \vdots\\ x_{N-1}\end{bmatrix}$, and transform it as shown in (7):

$$X_f=H\times X \tag{7}$$

Afterwards, the initial point is extracted using the k-means algorithm, and the point is selected based on amplitude weighting, as shown in formulas (8) and (9):

$$A_n=\sqrt[2]{\mathrm{Re}(X(n))^2+\mathrm{Im}(X(n))^2} \tag{8}$$

$$P_n=\arctan(\mathrm{Im}(X(n))/\mathrm{Re}(X(n))) \tag{9}$$

Among them, the amplitude is represented as A_n, phase is represented as P_n, and its original data is represented as $X(n)$.

The calculation formula for the central phase point of each data is shown in (10):

$$\begin{cases} \rho_n = \sum_j^N w_n * P_n \\ w_n = A_n / \sum_j A_j \end{cases} \tag{10}$$

The distance calculation formula between this sample and other samples in the same category is shown in (11):

$$d_{ns_n} = \sum_{j \in c_n} \sqrt[2]{\left\| x_n^{ap} - x_j^{ap} \right\|} \tag{11}$$

Among them, the central phase point data is divided and represented as $c_n, n \in [0, K]$, with a division quantity of K.

Then, use the roulette wheel method to randomly select the initial centroid $\mu_1, \mu_2, \ldots \mu_k \in X_f$, and repeat the above process until the classification is complete.

In order to test the performance of the module, this article selected three common fitness movements for quality evaluation accuracy testing, and the results are shown in Figure 2. Action 1 in the result is squatting against the wall, and its judgment criterion point is the parallelism between the thighs and the ground; Action 2 is a flat support, with two action judgment points; Action 3 is a Y-shaped review extension, and the criterion for judging its action is the angle between the body and the ground. The judgment results show that the module is capable of collecting and analyzing data on different fitness movements, with an accuracy rate of over 93%, and even some movements have an accuracy rate of over 97%. This indicates that the module can accurately obtain and analyze corresponding data from fitness enthusiasts, providing basic data support for optimizing fitness effects in the future.

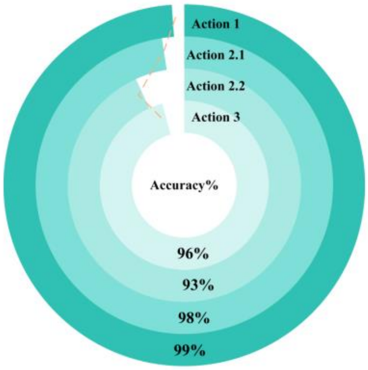

Figure 2: Analysis results of quality data for three common fitness exercises.

As shown in Figure 3, the module compares the average values of the style classification indicators for sports data, which also includes three other classification methods for comparison. This test includes two different datasets, with Dataset A having a much larger amount of data than Dataset B. In dataset A, the performance indicators of the module in this article perform the best, significantly higher than the other three classification methods; In dataset B, the performance of the module in this article is still good, but slightly weaker than in dataset A, and higher than the performance indicators of the other three methods. From this, it can be seen that the module classification performance in this article has good generalization ability, can adapt to the variability of data, and performs better in complex motion data analysis, which can adapt to practical application environments.

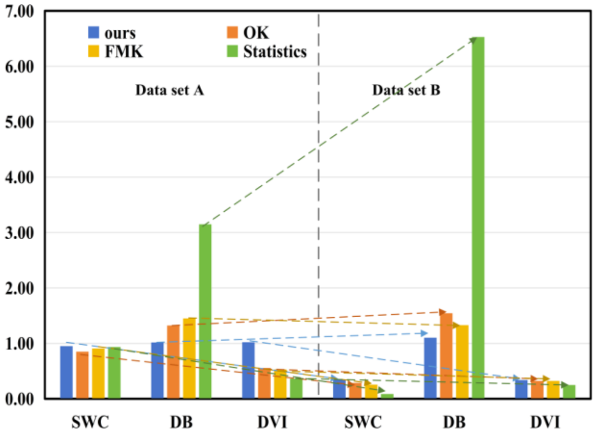

Figure 3: Comparison of average index values for style classification of sports data in the module.

3.2 Quantitative Evaluation Model for Fitness Effects Based on Xgboost Algorithm

The analysis of exercise data is not only to help fitness enthusiasts understand their fitness status from multiple perspectives and angles but also to provide decision-making data for optimizing fitness plans and effects in the future. Therefore, to improve the performance of fitness optimization strategies, this article adopts the XGBoost algorithm to achieve a quantitative evaluation of fitness effects, allowing fitness enthusiasts to see that their fitness has reached the desired state intuitively. In the process of quantification, it is necessary first to determine the corresponding exercise indicators based on the characteristics of different fitness programs and then calculate the corresponding exercise effect prediction values according to the XGBoost algorithm model, as shown in formula (12):

$$Y_n = M(X_n^1, X_n^2, ..., X_n^{15}) \tag{12}$$

Among them, the algorithm model is represented as M .

The number of sample datasets is N , the corresponding input variables are denoted as X_n , and each variable has fifteen features. The sample corresponding variable is denoted as y_n Assuming the number of model trees is G , the predicted value is shown in formula (13):

$$\begin{cases} \hat{y}_n = \sum_{g=1}^{G} f_g(X_n) \\ F = \{f(X) = \mu_l(X)\} \end{cases} \tag{13}$$

In the formula, $f_g \in F$, the regression tree set space is represented as F , the regression tree is represented as $f(X)$, and the leaf weights in the tree with g sequences in the sample are represented as $f_g(X_n)$. The number of sequences is represented as X, the leaf nodes of the samples are represented as $l(X)$, and the corresponding scores are represented as μ.

When the number of iterations is t, the predicted result is shown in (14):

$$y = \hat{y}_n^{t-1} + f_t(X_n) \tag{14}$$

The XGBoost objective function obtained by simplifying the second-order Taylor expansion is shown in (15):

$$\begin{cases} J(f_t) = \sum_{n=1}^{N}[k_n w_l(X_n) + \frac{1}{2}h_n w2_l(X_n) + \beta T + \delta\frac{1}{2}\sum_{j=1}^{T} w_j 2 \\ k_n = \frac{\partial L(y_n, \hat{y}^{t-1})}{\partial \hat{y}^{t-1}} \\ h_n = \frac{\partial 2L(y_n, \hat{y}^{t-1})}{\partial \hat{y}^{t-1}} \end{cases} \tag{15}$$

Among them, the loss function is represented as L, the total number of leaf nodes is represented as T, the regularization term of leaf nodes with a score of $L2$ is represented as $w_j 2$, and the penalty coefficient is represented as β.

In order to enable fitness enthusiasts to intuitively and effectively obtain analysis results, this article also uses SHAP to explain the importance of the characteristics that affect fitness performance. At the same time, it can also achieve visualization of exercise data and fitness performance analysis results. SHAP is a method used to explain machine learning models, based on the Shapley value theory in game theory. The Shapley value was originally used to solve the allocation problem of participants' contribution to the total revenue in cooperative games, but in the SHAP algorithm, it is used to measure the contribution of each feature in the machine learning model to the model output.

Assuming the input in the model is represented as its corresponding predicted output value $v(N)$, the contribution calculation formula for each feature is shown in (16):

$$\psi_n = \sum_{S \in N\{n\}} \frac{\left|S\right|!(n - \left|S\right| - 1)!}{n!}[v(S \cup \{n\} - v(S)] \tag{16}$$

In order to test the performance of the module, this article selected running as the testing item, and Table 1 shows the characteristics of running fitness indicators.

Index	*Meaning*	*Index*	*Meaning*
X1	Minimum heart rate within heart rate range 1 (bpm)	X9	Average speed (m/min)
X2	Minimum heart rate within heart rate range 2 (bpm)	X10	Maximum speed (m/min)
X3	Lowest heart rate within heart rate range 3 (bpm)	X11	Calories burned (kcal)
X4	Minimum heart rate within heart rate range 4 (bpm)	X12	Hydration (fl. oz)
X5	The lowest heart rate within the heart rate range of 5 (bpm)	X13	Duration of exercise (min)
X6	Maximum heart rate (bpm)	X14	Total elevation increase (foot)
X7	Resting heart rate (bpm)	X15	Total decrease in altitude (foot)
X8	Movement distance (m)	X16	Running power value

Table 1: Characteristics of Running Fitness Exercise Indicators.

Considering the differences in athletic performance and characteristics between male and female fitness enthusiasts, quantitative evaluations were conducted on the effects of middle and long-distance running for both genders. The results of the module performance test are shown in Figure 4. The results showed that both male and female fitness enthusiasts were able to impress with an accuracy rate of over 85% in the evaluation of middle and long-distance running, and had a good recall and F1 performance. This indicates that the module can effectively quantify the fitness effects of fitness enthusiasts of different genders and support application experiments.

Figure 4: Module performance test results.

4 EXPERIMENTAL RESULTS

In the application experiment, this article selects running fitness projects as the application testing objects, with an equal number of males and females. 70% of the data is used as the training set for the model, and 30% is used as the testing set for the model. As shown in Figure 5, the relationship between exercise centre rate and speed, as well as speed and strength, of randomly selected fitness participants in the test set who are male and engage in long-distance running. The results in Figure (a) show that the exerciser's heart rate increases significantly when the speed is low. As the speed decreases, the heart rate decreases and remains relatively stable. After that, increasing the speed also leads to a significant increase in heart rate. The results in Figure (b) show that the fitness practitioner's speed is relatively stable within the range of 1.5m/s to 2.2m/s, and their strength increases significantly beyond this range. Based on the two results, it can be seen that the fitness enthusiast can maintain a relatively stable heart rate and strength output when running at a speed of around 2.0m/s.

As shown in Figure 6, the analysis results between the heart rate and exercise duration of the fitness participant indicate that the model proposed in this paper can effectively classify and visually present the relevant data of the fitness participant. Fitness enthusiasts can understand the similarities and differences between their body and weight based on the centroid position marked in the diagram. To help fitness enthusiasts have a more comprehensive and multidimensional understanding of the gap between their own fitness results and those of others, the model also conducted running feature importance statistics based on the analysis results of the exercise data of the test subjects participating in the application experiment, allowing the fitness enthusiast to compare the corresponding data more intuitively.

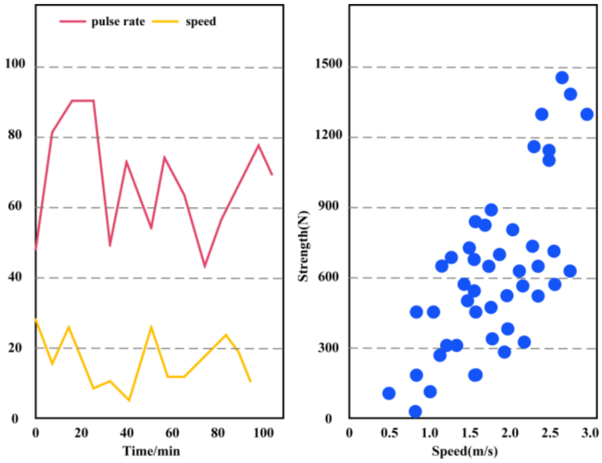

Figure 5: Relationship between exercise centre rate and speed, speed, and strength of random fitness enthusiasts.

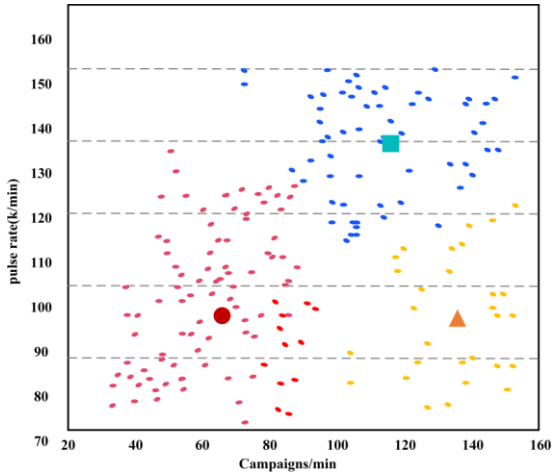

Figure 6: Analysis results between heart rate and exercise duration of fitness enthusiasts.

As shown in Figure 7, the overall SHAP values of female and male long-distance running characteristics are presented. The colours of different data in the graph represent the size of the corresponding values, and the order from dark to light indicates the value from large to small. As the test target is males, the SHAP value of male long-distance running characteristics is mainly analyzed. The results show that when men engage in fixed-distance running exercises if the average exercise speed is not high, they cannot achieve the desired exercise effect and may even affect the exercise effect. If the average speed is appropriately increased, the sports effect can be significantly improved, but this effect will tend to stabilize when the average speed reaches a certain value. Due to the certain correlation between exercise effectiveness and the heart rate of fitness enthusiasts, appropriately increasing heart rate can effectively improve the exercise effectiveness of fitness enthusiasts. Therefore, this conclusion is consistent with the above analysis of exercise data, and fitness enthusiast should increase their exercise speed appropriately during running to achieve the goal of increasing heart rate.

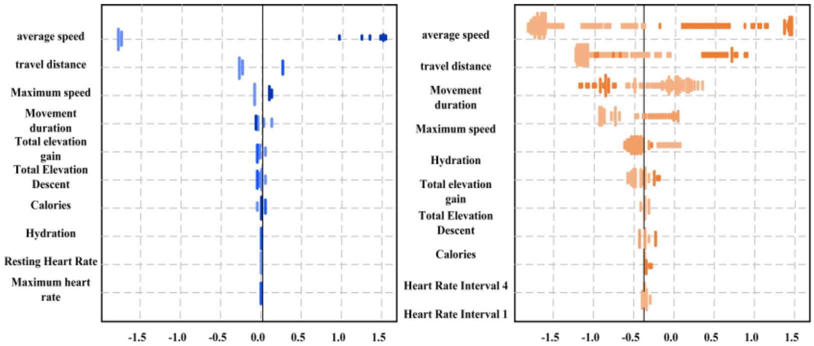

Figure 7: Overall SHAP values of female and male long-distance running characteristics.

In addition, the running posture of fitness enthusiasts has a significant impact on exercise effectiveness. Correct exercise posture can not only improve exercise effectiveness and running speed but also effectively reduce sports injuries. As shown in Figure 8, the fitness enthusiast adjusted their running posture based on the analysis of historical exercise data. This indicates that the model presented in this article can assist fitness enthusiasts in optimizing and adjusting their exercise posture based on the analysis of their exercise data and posture recognition, thereby improving exercise effectiveness.

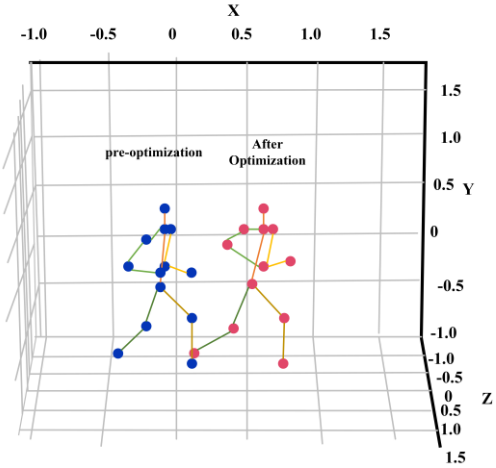

Figure 8: The fitness enthusiast's adjustment of running posture based on historical exercise data analysis results.

5 CONCLUSIONS

The improvement of sports data analysis ability is an inevitable trend in the development of the fitness industry. It can not only help fitness enthusiasts understand their own sports status from all aspects but also effectively help them adjust their exercise plans and postures and improve exercise effectiveness. Compared with traditional sports data analysis methods, computer-aided technology can improve analysis efficiency and accuracy, with the ability to process massive and complex data, and can enhance the real-time, integrity, and effectiveness of data analysis. Therefore, this article combines machine learning algorithms to construct a model for analyzing sports data and quantitatively evaluating fitness effects. By using a monocular camera, data collection errors are reduced and fitness posture recognition rates are improved. To improve the performance of sports data analysis, this article combines the k-means algorithm to implement sports data style classification and more accurately analyzes the corresponding data based on the sports styles of different fitness enthusiasts. In addition, this article also combines the XGBoost algorithm and SHAP algorithm to quantify the exercise effects of fitness enthusiasts and visualize the analysis results. The experimental results show that the model in this article has good data collection ability and motion posture quality evaluation ability, and the evaluation accuracy of different movements can reach over 93%. It can demonstrate high model performance and good stability in different datasets. In application experiments, this model can effectively analyze the relevant exercise data of fitness enthusiasts, presenting the correlation between important exercise features and other features from multiple perspectives, helping fitness enthusiasts to comprehensively understand their own exercise status. At the same time, the model can also display the overall SHAP values of male and female characteristics in the same exercise based on existing analysis object data, helping fitness enthusiasts to conduct horizontal data analysis and comparison. Based on the analysis results, this model can help fitness enthusiasts adjust their exercise posture and plan, achieving the goal of improving exercise effectiveness.

6 ACKNOWLEDGEMENT

Sports science and technology research projects, Research on the promotion mechanism of deep integration of national fitness and national health in Hebei Province, 2024QT02.

Qianqi Zhao, https://orcid.org/0009-0006-1307-350X
Zhenzhen Bie, https://orcid.org/0009-0005-9875-3493
Yang Li, https://orcid.org/0000-0002-9397-3906

REFERENCES

[1] Alatas, B.: Sports inspired computational intelligence algorithms for global optimization, Artificial Intelligence Review, 52(1), 2019, 1579-1627. https://doi.org/10.1007/s10462-017-9587-x
[2] Chen, X.; Kasgari, A.-T.-Z.; Saad, W.: Deep learning for content-based personalized viewport prediction of 360-degree VR videos, IEEE Networking Letters, 2(2), 2020, 81-84. https://doi.org/10.1109/LNET.2020.2977124
[3] Cong, C.; Fu, D.: An AI-based research on optimization of university sports information service, Journal of Intelligent & Fuzzy Systems, 40(2), 2021, 3313-3324. https://doi.org/10.3233/JIFS-189371
[4] Dubey, S.; Dixit, M.: A comprehensive survey on human pose estimation Approaches, Multimedia Systems, 29(1), 2023, 167-195. https://doi.org/10.1007/s00530-022-00980-0
[5] Eden, J.; Bräcklein, M.; Ibáñez, J.; Barsakcioglu, D.-Y.; Pino, G.; Farina, D.; Mehring, C.: Principles of human movement augmentation and the challenges in making it a reality, Nature Communications, 13(1), 2022, 1345. https://doi.org/10.1038/s41467-022-28725-7
[6] Kulkarni, S.; Deshmukh, S.; Fernandes, F.; Patil, A.; Jabade, V.: Poseanalyser: A survey on human pose estimation, SN Computer Science, 4(2), 2023, 136. https://doi.org/10.1007/s42979-022-01567-2
[7] Liu, J.; Wang, L.; Zhou, H.: The application of human–computer interaction technology fused with artificial intelligence in sports moving target detection education for college athlete, Frontiers in Psychology, 12(1), 2021, 677590. https://doi.org/10.3389/fpsyg.2021.677590

[8] Ma, N.; Wu, Z.; Cheung, Y.-M.; Guo, Y.; Gao, Y.; Li, J.; Jiang, B.: A survey of human action recognition and posture prediction, Tsinghua Science and Technology, 27(6), 2022, 973-1001. https://doi.org/10.26599/TST.2021.9010068

[9] Marín, J.; Blanco, T.; Torre, J.; Marín, J.-J.: Gait analysis in a box: a system based on magnetometer-free IMUs or clusters of optical markers with automatic event detection, Sensors, 20(12), 2020, 3338. https://doi.org/10.3390/s20123338

[10] Menolotto, M.; Komaris, D.-S.; Tedesco, S.; O'Flynn, B.; Walsh, M.: Motion capture technology in industrial applications: A systematic review, Sensors, 20(19), 2020, 5687. https://doi.org/10.3390/s20195687

[11] Muzammal, M.; Talat, R.; Sodhro, A.-H.; Pirbhulal, S.: A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks, Information Fusion, 53(1), 2020, 155-164. https://doi.org/10.1016/j.inffus.2019.06.021

[12] Ryselis, K.; Petkus, T.; Blažauskas, T.; Maskeliūnas, R.; Damaševičius, R.: Multiple Kinect based system to monitor and analyze key performance indicators of physical training, Human-Centric Computing and Information Sciences, 10(1), 2020, 1-22. https://doi.org/10.1186/s13673-020-00256-4

[13] Salloum, G.; Tekli, J.: Automated and personalized nutrition health assessment, recommendation, and progress evaluation using fuzzy reasoning, International Journal of Human-Computer Studies, 151(1), 2021, 102610. https://doi.org/10.1016/j.ijhcs.2021.102610

[14] Santos, T.-M.; Barroso, M.-F.; Ricco, R.-A.; Nepomuceno, E.-G.; Alvarenga, É.-L.; Penoni, Á.-C.; Santos, A.-F.: A low-cost wireless system of inertial sensors to postural analysis during human movement, Measurement, 148(1), 2019, 106933. https://doi.org/10.1016/j.measurement.2019.106933

[15] Singh, T.; Vishwakarma, D.-K.: A deeply coupled ConvNet for human activity recognition using dynamic and RGB images, Neural Computing and Applications, 33(1), 2021, 469-485. https://doi.org/10.1007/s00521-020-05018-y