J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 5941--5961

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

Combinatorial Press

Research on Modeling and Pattern Recognition Algorithm for Youth Sports Training Data in Internet of Things Environment

Xuemin Han[1], Peng Guo[2,✉], Ziqi Deng[1], Xu Han[3], Hong Wang[1]

[1] *Hainan University, Danzhou, Hainan, 573717, China*

[2] *Cangzhou Transport University, Huanghua, Hebei, 061199, China*

[3] *School of International Education, Henan University, Zhengzhou, Henan, 450000, China*

ABSTRACT

The aim of this paper is to construct a data model applicable to youth sports training in the IoT environment and develop efficient pattern recognition algorithms to achieve accurate analysis and assessment of youth sports training status. The features of youth sports training data collected by IoT technology are extracted through a combination of deep learning and feature decomposition. The feature vectors obtained from feature extraction are inputted into the Long Short-Term Memory (LSTM) network to generate the data model of youth sports training in this paper and predict the state of youth sports training. The prediction results are input as features into the Support Vector Machine (SVM) algorithm, and these features are extracted using the Empirical Modal Decomposition (EEMD) method, and at the same time, the hierarchical idea is utilized to realize the recognition of youth sports training patterns. The results of the study showed that the errors of the results of predicting youth sports training states using the LSTM model were mostly within 0 ± 0.5. The prediction accuracies of the model on the test set for the three athletic training state metrics were 96.80%, 99.40%, and 98.80%, respectively. Meanwhile, the performance of the SVM model for youth athletic training state pattern recognition using the SVM model was significantly superior, with 100% accuracy on the test set for four models, including pattern 2.

Keywords: LSTM; SVM; EEMD; deep learning; pattern recognition; youth sports training

1. Introduction

Teenagers are a diversified group with infinite possibilities in them, and the content of their learned knowledge is stable within a certain period of time, which can further promote the comprehensive

✉ Corresponding author.
E-mail address: hainanxuemin@126.com (P. Guo).

Received 12 January 2024; Revised 05 March 2024; Accepted 25 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-326

shaping of teenagers. With the development of modern science and technology, more and more teenagers are addicted to online games, web gambling, cell phone socialization and other virtual worlds, and lack of physical exercise and sports training, which makes the physical quality gradually decline, and the health condition is getting worse and worse, and in serious cases, even threaten the life safety [1-4]. At the same time, with the development of society and the improvement of living standards, adolescents are facing more and more pressure of study and life, which leads to the decline of adolescents' physical quality, and physiological and psychological diseases follow [5-6]. Adolescents are in the period of growth and development, and exercise training under scientific guidance can not only promote bone growth and development, enhance muscle strength and endurance, but also promote the improvement of cardiorespiratory fitness, improve body adaptability, and enhance motor skills and coordination, thus enhancing physical fitness, improving immunity, and preventing diseases [7-10]. Youth sports training also helps to regulate emotions, release stress, promote mental health level, improve concentration and learning efficiency, develop teamwork and self-confidence, and shape sunny and healthy personality traits [11-12]. Youth sports training is crucial for their physical health and overall development, and deserves attention and concern. However, the current youth sports training lacks scientific, systematic, comprehensive and individualized. Many training programs lack scientific basis, just mechanical repetition of certain training actions, the lack of systematic planning and guidance, and the existence of a set of training programs used by all groups of people, completely ignoring the youth's individual needs and training goals, which can not effectively improve the physical fitness level of youth, and may even cause damage to their bodies [13-15]. Furthermore, over-training is also a hidden danger in youth sports training. Some adolescents may over-train in order to pursue better training results, resulting in excessive load on the body, increasing the risk of injury, and long-term over-training may also adversely affect the physical and mental health of adolescents [16-17]. At the same time, there is a lack of proper identification of training injuries during the training process, which makes it difficult to provide early warning of injuries, and the warning mechanism is backward [18]. In addition, the lack of motivation and interest in sports among adolescents is also a key issue that affects the effectiveness of sports training. With the fast-paced life and diversification of entertainment in modern society, adolescents are more inclined to choose other forms of recreational activities to maintain the basic needs of the body, and have little interest in sports training and lack of sustained motivation [19-20]. Therefore, about how to carry out scientific sports training is one of the problems that need to be solved in the current society.

Nowadays, with the rapid development of science and technology, the Internet of Things has become an inevitable trend. And the interconnectivity and intelligence of IoT devices make the modern society present a brand-new model, through which the IoT can monitor, collect and analyze the sports training data, so as to provide professional services [21-22]. In sports training, intelligent monitoring of training status requires miniaturization of equipment, not only to save space and lightweight and flexible, but also conducive to reducing energy consumption, improving energy use efficiency and achieving sustainable development [23-24]. Based on this, the Internet of Things (IoT) technology has been widely used in various types of sports training, all based on sports monitoring to identify and prevent sports injuries, and enhance sports training, etc. Dan et al [25] designed a sports training model supported by Internet of Things (IoT) technology and computer multimedia technology for rapid acquisition of postural and positional data in human sports training, and by aggregating the action data from sports training, a high level of visual effect is obtained by processing the data. processing to obtain a high-level visual effect display, thus enhancing the training effect.Zhang et al [26] designed a wearable form of continuous monitoring system for individual sports training by combining IoT technology and artificial intelligence, with which the effect of sports training was enhanced.Zhao et al [27] constructed a sports training recognition system to analyze the behavior of IoT video with a dynamic temporal regularization model for the spatial description of human joints, which can realize frame-by-frame recognition and does not have individual differences.Zhen et al [28] relied on IoT and used compact wearable devices and microprocessors to construct an intelligent Jiankang monitoring system that can monitor and track body parameters and reaction time under stimulation, which provides the possibility of early warning of sports injuries.Wilkerson et al [29] identified and analyzed the physical function data and the effects of past injuries before entering sports through IoT and analytics to prevent subsequent sports injuries and reduce the risk of sports injuries.Zhao and You [30] introduced IoT technology to develop a wearable sports posture measurement system for sports training, which can monitor and analyze the body data received by sensing online to facilitate coaches' development of training program. Based on the problems in youth sports training and the prominent role of IoT technology in sports training, there is a current need to consider how IoT technology can be better applied to the field of youth sports training to improve physical fitness.

In this paper, with the advantage of IoT technology, a monitoring method integrating wireless sensor nodes and Sink nodes is constructed to acquire youth sports training data in real time. For the acquired training data, on the basis of determining its information distribution, deep learning and feature decomposition methods are used to comprehensively mine and parse the training data. The AT critical value point anaerobic threshold is introduced to construct an exercise training model with fusion rules, which provides support for the data modeling of adolescent exercise training. On the basis of fully realizing data acquisition and feature extraction, this paper uses the long short-term memory network (LSTM) to construct an adolescent exercise training data model to predict the adolescent exercise training status. The sports training state prediction results are used as the features, and the support vector machine (SVM) algorithm is combined to generate the adolescent sports training pattern recognition model. Fourteen adolescents were selected as the experimental subjects of this paper, and their sports training data were collected using the method proposed in this paper, and the validity of the data model and the pattern recognition model in this paper were examined separately, so as to provide strong support for scientific guidance of adolescent sports training.

2. Athletic training data collection and feature extraction

2.1. Data collection

In this paper, with the advantage of IoT technology and the self-organizing ability of wireless sensor networks, following the IEEE 802.15.4 standard, we aim to construct a software and hardware validated monitoring method integrating wireless sensor nodes and Sink nodes to acquire training data in real time, as shown in Fig. 1. The core components of the method in this paper include a miniature wireless sensor front-end worn by the adolescents, an RFID identifier, an IoT gateway node and a central server. Each adolescent is equipped with a unique RFID tag, enabling the system to track and monitor the physiological status of multiple individuals simultaneously. The server-side Sink node receives and synthesizes the massive amount of data from the wireless sensor network. Subsequently, the data is sequentially transferred to the connected computers via an RS 232 serial control chip.

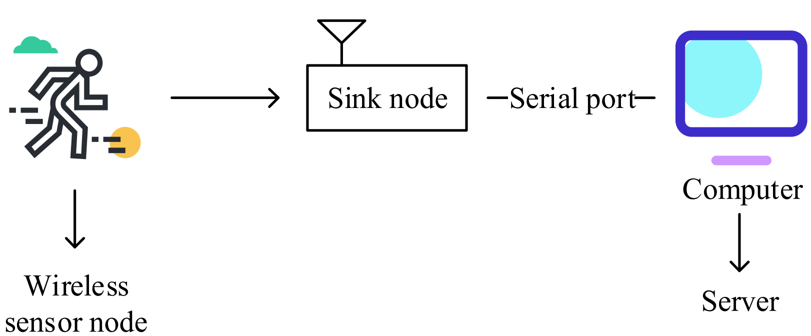


Fig. 1. Schematic diagram of collecting training data

2.2. Feature extraction

2.2.1. Determining the information distribution of athletic training data. In this paper, the focus of feature extraction of youth sports training data is centered on constructing a joint probability distribution model of constraint and explanatory variables in sports training load, aiming to comprehensively reveal the intrinsic laws of training load through this model. In order to understand the characteristics of sports training load more comprehensively, this paper introduces several important parameters, including resting heart rate, ventricular muscle contraction force, and wall thickness. The feature distribution of the sports training data is shown in Figure 2.

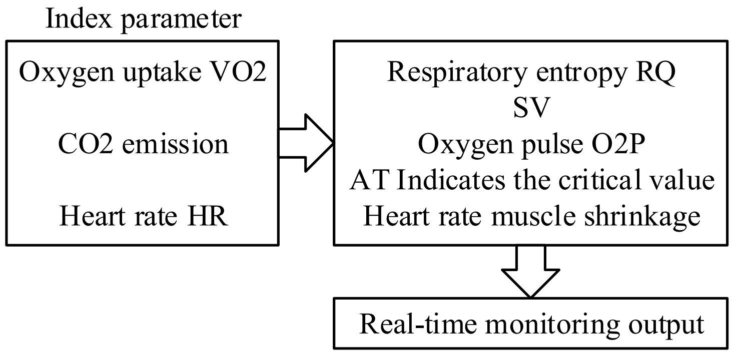


Fig. 2. Feature distribution of sports training data

Based on the characteristic distribution of data in Figure 2, considering the influence of factors such as oxygen uptake VO2, CO_2 excretion, and oxygen pulse O2P on cardiopulmonary fitness, we constructed a distribution model of influence constraints of cardiopulmonary fitness through comprehensive analysis and mathematical modeling:

$$RQ = \frac{VO_2}{SCO_2} \tag{1}$$

Where: SCO_2 is a multivariate fusion parameter, which is used to monitor the data of exercise training load in real time. With the help of principal component fusion technology and linear correlation decision-making strategy, the key component information indexes of sports training load can be quickly and accurately extracted and feature-matched. By this method, the matching set of youth sports training features is obtained:

$$S = \left\{ \frac{SCO_2 + 1}{OC}, \frac{SCO_2 + 2}{OC}, \cdots, \frac{SCO_2 + n}{OC} \right\} \tag{2}$$

Where: OC represents the joint distribution coefficient for real-time monitoring of sports training load data. In order to obtain a more accurate OC value, this paper analyzes the output constraints based on the maximum pulse output to ensure that the output sports training load data is both accurate and reliable.

2.2.2. Extracting sports training data features. In order to achieve more accurate data analysis and modeling, deep learning and feature decomposition methods are used to comprehensively mine and parse the training data to construct a data fusion model for sports training data:

$$B_j = \left[(k(t)+1), (k(t)+2), \cdots, (k(t)+n) \right] \tag{3}$$

Where: $k(t)$ is the characteristics of discrete samples. Using data fusion techniques, these characteristics are effectively integrated to form a comprehensive and accurate dataset, which is analyzed by regression analysis, and the results are shown in equation (4):

$$S = \sum_{i=0} \left[k(t) - \omega_i(t) \right]^2 + B_j \tag{4}$$

Where: $\omega_i(t)$ is the regression coefficient. The results of the quantitative analysis of lung function were fully incorporated into the process of data integration in order to model threshold management for exercise training data:

$$g_{ic} = S(\beta_1, \beta_2, \cdots, \beta_n) \tag{5}$$

Based on the in-depth profiling after data integration, the interrelated detection technique of sports training data is further developed. In order to assess the training status of adolescents more accurately, this paper introduces the AT threshold point anaerobic threshold, and by combining this indicator, an exercise training model with fusion rules is obtained:

$$E_2 = (g_{ic} + RQ)^2 \tag{6}$$

Real-time dynamic changes in athletic training data are captured through an efficient and accurate athletic training monitoring model.

3. Methodology for modeling athletic training data

Considering the multiplicity and complexity of youth sports training data, the idea of hierarchical modeling is adopted. The data model is divided into three layers: data layer, feature layer and model layer. The data layer is responsible for storing and managing the original collected data. The feature layer is responsible for storing and managing the original data. The feature layer extracts and selects features from the original data to obtain feature vectors that can effectively characterize the youth sports training status. The model layer constructs a mathematical model based on the data in the feature layer, which is used to analyze and predict the youth sports training status. Since the collection and feature extraction methods of youth sports training data have been discussed in the previous section, they will not be repeated in this section. Therefore, this section will focus on the use of Long Short-Term Memory (LSTM) network to predict and analyze the youth sports training data so as to achieve the data modeling effect.

3.1. LSTM model

Recurrent neural networks need to rely on recent information to predict the results, recurrent neural networks have a better performance. If the recurrent neural network encounters processing long time memory information, it needs to combine the information of the hidden layer at this moment, which will make the model computation and training time increase dramatically, in order to solve the problem of recurrent neural network in processing the information that needs to rely on a long time, based on the recurrent neural network, the long-short-term memory network (LSTM) model was proposed [31]. The LSTM model is a kind of special kind of recurrent neural network, which uses gated recurrent units instead of neurons in recurrent neural networks to realize the processing of information, and each unit contains one or more memory units, and the gate control mechanism can make the information pass selectively, avoiding the problem of large prediction error of long-term time information.

The long and short term memory network structure is composed of chained recurrent network structure. So it has a better ability to deal with time series than the recurrent neural network, and it can effectively solve the problem of long time dependence of the traditional recurrent neural network. Since the long and short-term memory network structure has been proposed, it has been continuously refined and improved, and its structure is similar to that of the recurrent neural network, which can avoid the problem of long-term dependence and has the ability to learn on its own. Using LSTM to predict the influence factors of youth sports training data can get better results.

The first step of the Long Short-Term Memory network model is to decide what information to let flow through the network, which is achieved by a forgetting gate. This structure can read and get a value between 0 and 1 to each network recurrent unit, the update formula of this structure is as in equation (7). If the currently predicted information is not related to the information that appeared before, the information at this point is discarded when the forgetting gate is used:

$$f_i = \sigma(W_j \cdot [h_{i-1}, x_i] + b_j) \tag{7}$$

The model then has to decide what information should be stored within the loop cell. This step is done mainly through the combination of the information input from the input gate and a new cell structure which contains the updated state. The structure generates a vector mainly through the sigmoid and tanh layers of the input layer. Then, these two parts are combined to perform an update of the cell's state. The updating equations for this process are shown in Eq. (8), Eq. (9) and Eq. (10):

$$i_i = \sigma(W_i \cdot [h_{i-1}, x_i] + b_i) \tag{8}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{9}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{10}$$

The previously obtained information is combined with the forgetting gate to eliminate irrelevant information. Then multiply the previously obtained useful information with the update gate to realize the process of updating the information. A sigmoid layer is run to determine which part will be output to the network structure. Eventually the part of the information that determines the output will be obtained. The output gate structure is updated as in equations (11) and (12):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

$$h_t = o_t * \tanh(C_t) \tag{12}$$

where t denotes the current information processing state of the network. b and c denote the bias vectors of the parameters. x_i denotes the input of state t. h_i denotes the hidden layer output of state t. o_i denotes the output layer input of state t. C_i is the hidden layer state. The loss function gives a clearer picture of the training process of the network.

3.2. Data modeling

The main use of youth sports training data modeling is to predict the state of youth sports training, the prediction is a temporal prediction problem, the establishment of long-short time and the sports training state model of the network using a many-to-one structure, set $\{y_0, y_1, \ldots, y_{m-1}\}$ for the state of sports training at various moments, $\{f_1, f_2, \ldots, f_m\}$ for the temporal information at various moments, $\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m\}$ for the prediction of the state of youth sports training at different moments of the use of long-short time memory network outputs $\hat{y}_m$ is the final model output. In this model, the input is the initial sports training state and time information as well as the current sports training state information to be predicted, and the output is the current sports training state prediction value. At moment t the input is the motion training state at moment $t-1$ and the time information at moment t, and the output is the predicted motion training state at moment t, and finally the last moment of the predicted motion training state is taken as the output. In this way, the output layer outputs the predicted data of the future time to predict the youth's sports training state.

4. Athletic training pattern recognition methodology

Athletic training pattern is the classification of youth athletic training, while athletic training pattern recognition refers to the process of classifying various athletic patterns by obtaining youth athletic training information. In this section, we study to achieve high precision youth sports training pattern recognition based on the output of data modeling in the previous section, combined with the support vector machine (SVM) algorithm.

4.1. Support vector machine algorithm

The support vector machine (SVM) algorithm is a kernel-based machine learning model for classification and regression tasks [32]. In recent years, the extraordinary generalization ability of support vector machines and their optimal solution and discriminative power have attracted the attention of the data mining, pattern recognition and machine learning communities. Support vector machines have been used as powerful tools for solving practical binary classification problems.

In order to further understand the principle of SVM in depth, it can be illustrated in conjunction with Figure 3. The solid and hollow dots in the figure represent two types of training samples, L is the classification line, L_1 and L_2 represent the samples closest to the classification line, and the three straight lines are parallel to each other. The distance between L_1 and L_2 is called the classification interval, which is represented by margin.

In order to get the optimal classification result, it is necessary to find a suitable location for the classification line L so that the margin reaches its maximum value.

Assuming that the coordinates of all the samples are the general form of the linear discriminant function in $(x_i, y_i), i = 1, 2, 3, \cdots, n, d$ -dimensional space is $f(x) = kx + b$, and the equation of the classification line L is:

$$kx + b = 0 \tag{13}$$

Normalize the discriminant function $f(x)$, which is satisfied by all samples being classified:

$$| f(x) | > 1 \tag{14}$$

At this time, the classification interval is equal to $2/\|k\|$. According to the definition of the optimal classification surface, when $2/\|k\|$ reaches the maximum value, we get the optimal classification surface, and at this time, $\|k\|$ reaches the minimum value, then the classification surface that satisfies $\frac{1}{2}\|k\|$ (or $\frac{1}{2}\|k\|^2$) is the optimal classification surface, and the training samples on L_1, L_2 are the support vectors.

Based on the above analysis, the solution of the optimal classification surface can be expressed as the following constraint function solution, namely:

$$\varphi(k) = \frac{1}{2}\|k\|^2 \tag{15}$$

It can also be expressed as the solution of a Lagrangian constrained problem:

$$\Lambda(k, \beta, \alpha) = \frac{1}{2}\|k\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i(kx_i + \beta) - 1 \right] \tag{16}$$

By obtaining the minimal value of A by partial derivation of k, β in the above equation, a definite pattern of human motion can be obtained from the sensor data.

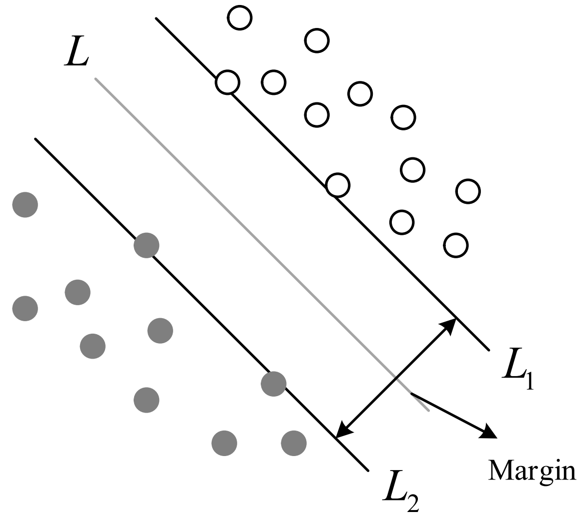


Fig. 3. Principle of optimal classification surface

4.2. Feature extraction

The feature extraction in this section is different from the feature extraction after the collection of youth sports data, which is mainly used to extract the features of youth sports training states output with data modeling, i.e., the sports training pattern recognition in this section inputs the predicted outputs of youth sports training states from the LSTM model in the previous section as features into the pattern recognition model in this section. Considering that the features of the predicted output youth sports training state data are generally nonlinear and nonsmooth signals, the features of these data are extracted using empirical modal decomposition (EEMD) [33]. The feature extraction process is as follows:

Input the LSTM sports training state prediction results into the pattern recognition model. Segment the input data into waist nodes and thigh area nodes. The waist node is close to the center of gravity of the human body and is mainly used to detect the overall trend of adolescent sports training. The thigh area node is used to detect the details of sports training, such as the speed of movement and the angle of movement. Each node can obtain acceleration information $(\lambda_x, \lambda_y, \lambda_z)$ and angular velocity information (ξ_x, ξ_y, ξ_z) in three axial directions at the same time.

For the motion data in three axial directions, the nodes can measure the inertial signal components in x, y, and z directions, which are used to express the motion of the adolescent in different directions. The combined acceleration and combined angular velocity are used to express the overall motion of the limb and are calculated as follows:

$$\lambda = \sqrt{\lambda_x^2 + \lambda_y^2 + \lambda_z^2} \tag{17}$$

$$\xi = \sqrt{\xi_x^2 + \xi_y^2 + \xi_z^2} \tag{18}$$

4.3. Layering algorithms

The pattern objects detected in this design mainly focus on the basic movement patterns in youth sports training, such as reaction coordination training, core strengthening training, continuous aerobic training, dynamic balance training, and dynamic stretching training. In order to further improve the efficiency of extracting features from the sports signals, the signals can be classified in advance according to their characteristics. Firstly, the signals can be categorized into six patterns such as sensitive coordination class, strength development class, etc. Sensitive coordination class of exercise training such as change of direction movement training, reaction coordination training. Strength development category of exercise training such as self-weighted strength training, core strengthening training and so on. On this basis, the posture of the movement pattern is analyzed by combining the calculation formula such as angular velocity, so that the movement training pattern can be obtained more quickly.

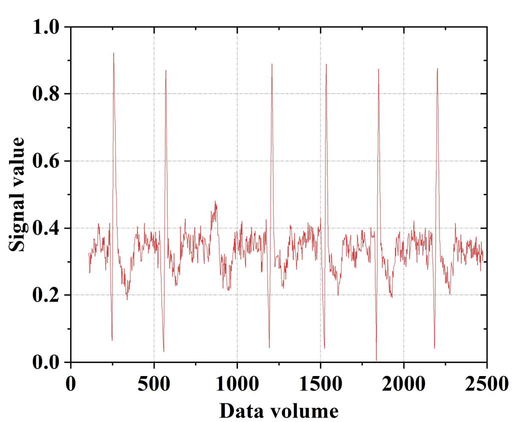
5. Experimental results and analysis

5.1. Data modeling analysis

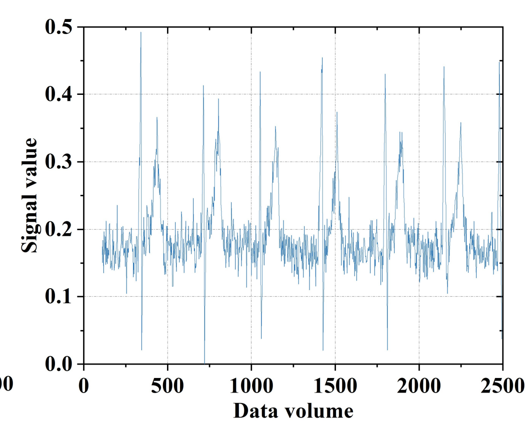
5.1.1. Data acquisition and parameterization. In this section, data collection method based on IoT technology is used to obtain adolescent sports training data signals. In order to obtain more accurate adolescent sports training data, this experiment collects the sports training data of 14 adolescents, including 7 boys and 7 girls. The age of the collectors was controlled from 16 to 18 years old, and all of their athletic training conditions were normal. The height range of the males was between 1.60 and 1.80m, and the height range of the females was between 1.55 and 1.75m earlier. The acquisition frequency of the sensors was set at 80Hz, and the data were collected from five aspects of the adolescents' physiological function, athletic performance, physical condition assessment, environment and psychology, and long-term development, and 2500 combinations of frames of athletic data were obtained from each aspect. 500 sets of exercise data from each aspect were randomly selected as test data, and the rest were used as training data for implementing the data model. The data of this experiment was processed on Lenovo PC and MatlabR2018b was used as the software platform.

The LSTM neural network output layer has 3 nodes, i.e., it outputs the levels of fatigue, training efficiency, and injury risk (divided into 5 levels, denoted by L1-5) regarding the state of youth sports training. The input data, on the other hand, includes the collected data on five aspects: physiological function, athletic performance, physical state assessment, environment and psychology, and long-term development. After several optimizations to find the most appropriate neural network parameters to set, the number of hidden layers of the model was finally set to 4, the number of neurons within the hidden layers was set to 35, and the early stopping algorithm was set to prevent overfitting. After learning and training and testing, the prediction results of youth sports training status were obtained.

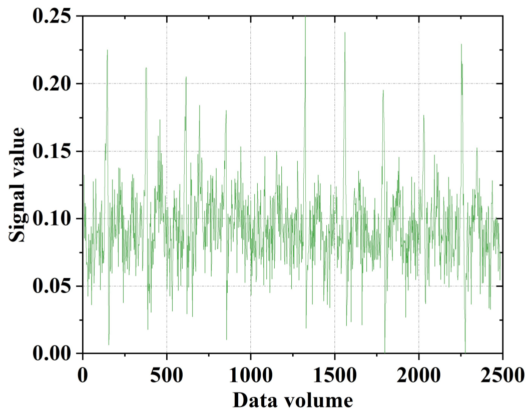
5.1.2. Feature extraction. The main work in this subsection is to extract the features of the collected data on physiological functions and other aspects using the feature extraction method proposed in this paper for youth sports training data. Figure 4 shows the raw signatures of the data from five aspects (denoted by D1-D5, respectively): physiological function, athletic performance, physical status, environment and psychology, and long-term development.



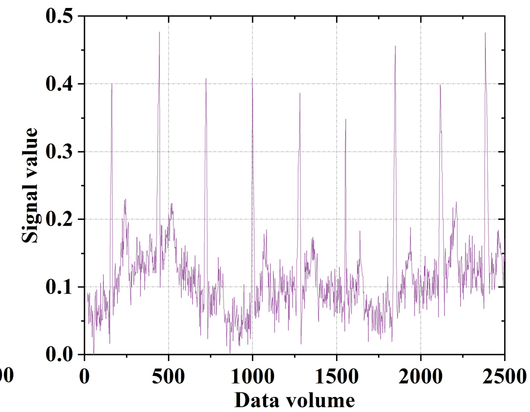
(a) D1



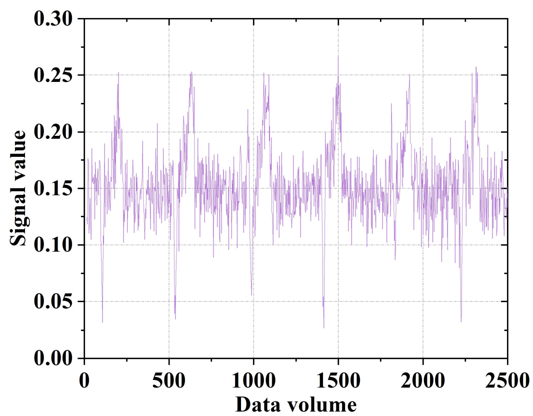
(b) D2



(c) D3



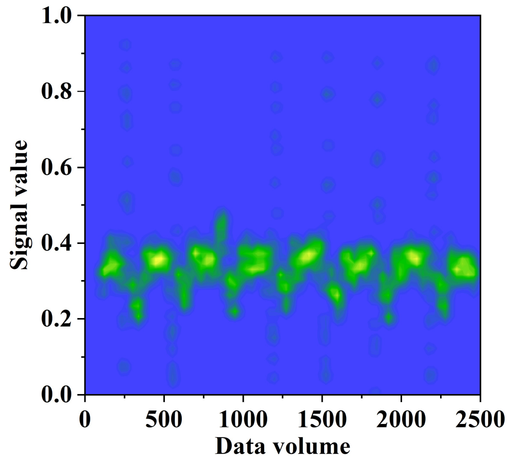
(d) D4



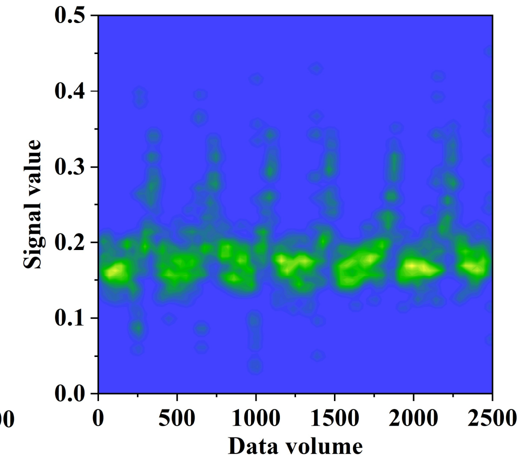
(e) D5

Fig. 4. Original motion signal data

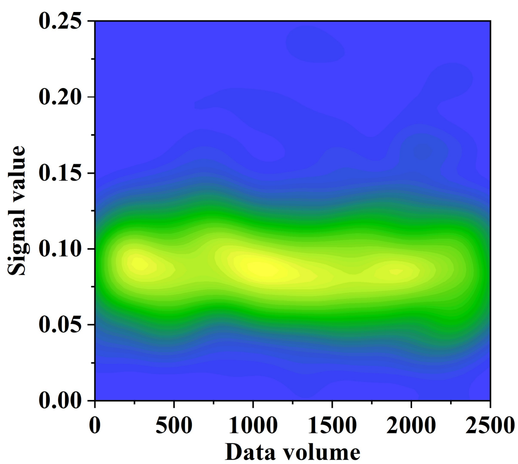
Based on the acquired youth sports training data signals, a suitable single sample length is first selected, and then feature extraction is performed using the combination of deep learning and feature decomposition proposed in this paper. The more data points a single sample contains, the richer the features it contains. However, if each sample contains too many data points, the total number of samples will be too small due to the limited data in the dataset, which is not conducive to the training of the LSTM model. By observing the pattern of the sea of stars of youth sports training data, it is found that the signal is periodic, and there are 300-800 data points between the peak and the peak of each cycle. Therefore, in order to ensure that each sample contains multiple complete cycles of sports signals, each sample is chosen to contain 1000 data points. The motion training signal features extracted by deep learning and feature decomposition are converted to 2D images in a uniform format, and this method maximizes the preservation of the features contained in the 1D signal. Using the method of this paper to propose the motion training signal features after converting to a 2D image is shown in Fig. 5, (a)-(e) denote the data categories such as physiological skills, which are denoted by D1-D5, respectively.



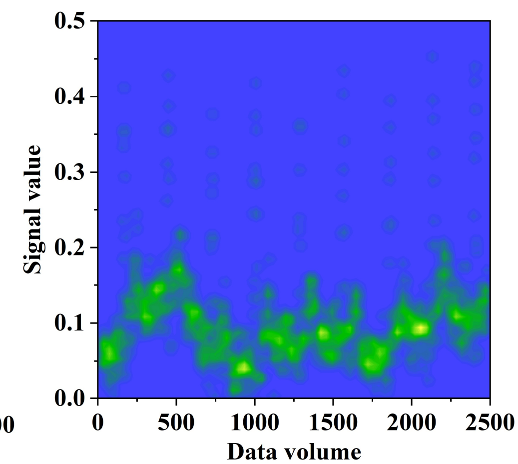
(a) D1



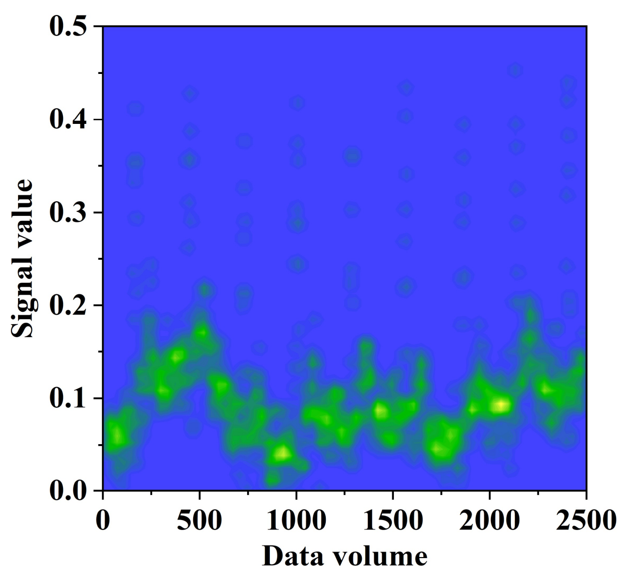
(b) D2



(c) D3



(d) D4

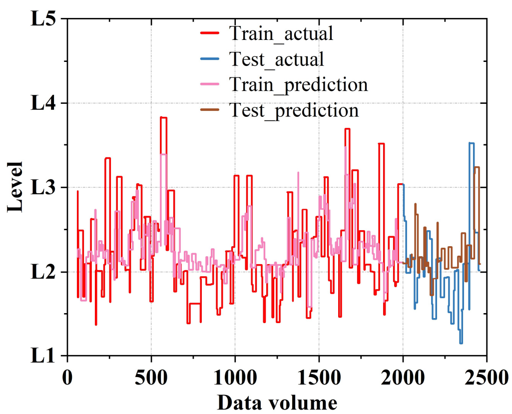


(e) D5

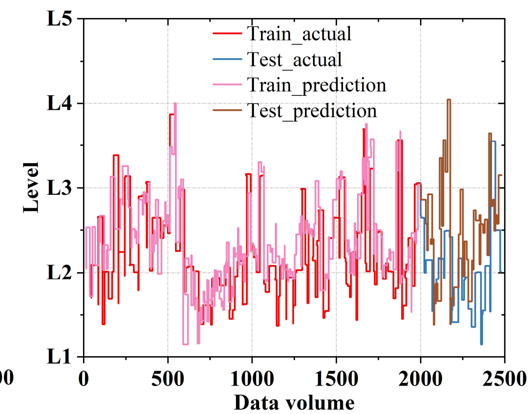
Fig. 5. Feature extraction results

5.1.3. Motion state prediction. Through data acquisition and feature extraction, this paper lays a solid foundation for the establishment of LSTM model. Physiological functions and other sports data are used as inputs to the LSTM model, and the data model for predicting the sports training status of adolescents is built through optimized training and learning. This section mainly explores the application effect of the LSTM prediction model based on the sports training data collected by IoT technology.

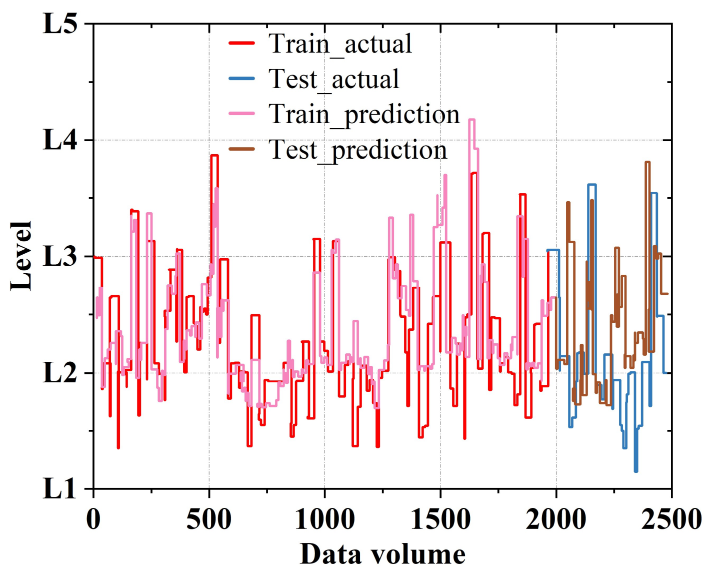
Figure 6 shows the results of the model on the whole dataset, and (a)-(c) indicate the output results of the indicators of sports training status such as fatigue level, training efficiency, and injury risk. The vertical coordinate in the figure is the rank of each sport training status indicator, and the horizontal coordinate is the amount of data in the whole dataset. It can be seen that in the prediction of youth sports training status, the model is more accurate in predicting the trend of the three sports training status indicators, especially the prediction of the training efficiency indicator fits better. Meanwhile, it can be seen from the distribution of actual and predicted values in the figure that the output levels of fatigue, training efficiency, and injury risk are concentrated in levels 1-4. Therefore, the follow-up study mainly focuses on levels 1-4 of each index.



(a) Degree of fatigue



(b) Training efficiency

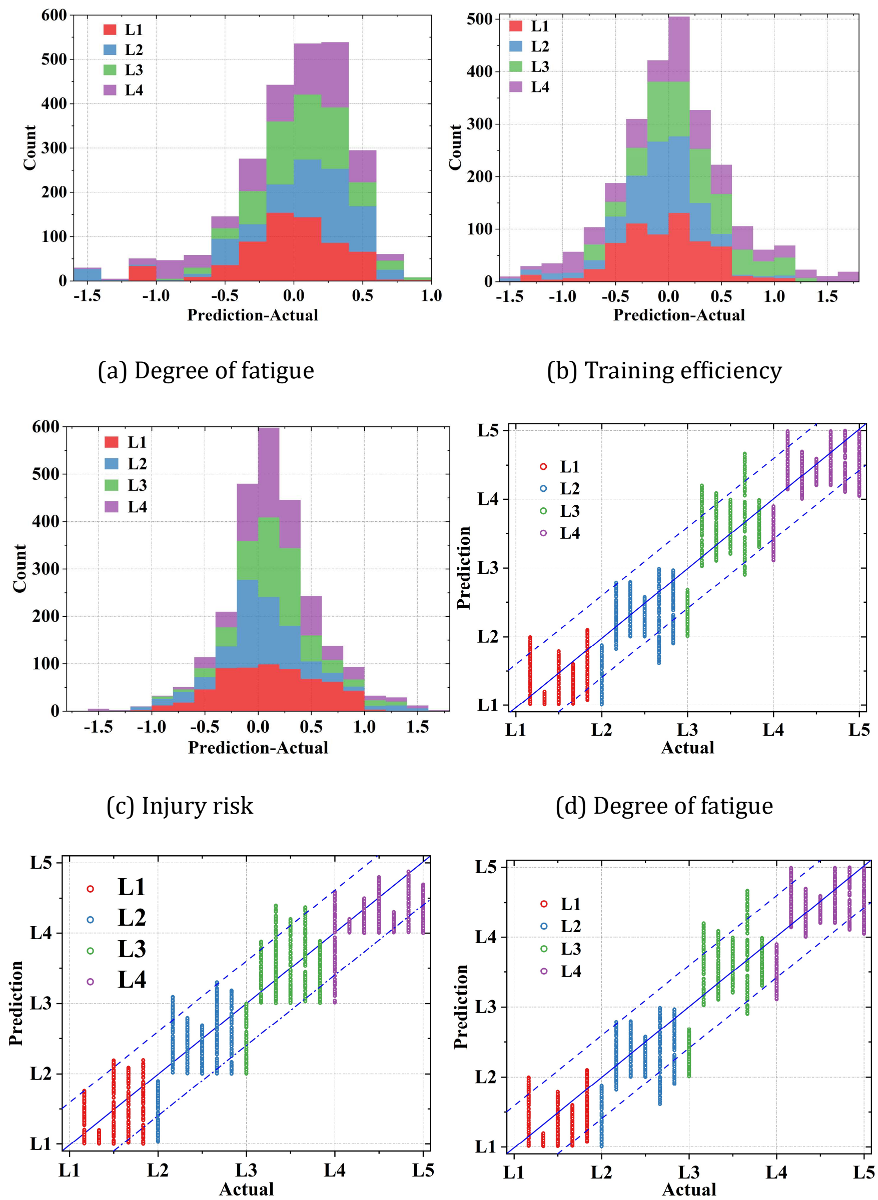


(c) Injury risk

Fig. 6. Scatterplot of prediction and actual values on the entire dataset

The horizontal axis of the histogram in Fig. 7 is the predicted index level of athletic training status minus the actual observed index level, which is plotted as a histogram of the error. From the histogram, it can be seen that most of the errors fall within 0 ± 0.5, indicating that the prediction results have a fairly high level of confidence. Moreover, the errors of the model for the prediction of training efficiency and injury risk are significantly smaller than the errors of the model for the prediction of fatigue, indicating that there is still room for optimization of the model for the prediction of fatigue, and efforts can be continued in this direction in the future.

The scatterplot in Figure 7 depicts the relationship between the predicted and observed indicator levels, and draws the auxiliary line where the indicator level is equal to the actual level (solid blue line), and the auxiliary line where the error is 0.5 (dashed blue line). It can be seen that the model predicts all three metrics better, and the LSTM model is more effective in predicting the training efficiency metrics compared to fatigue level and injury risk.



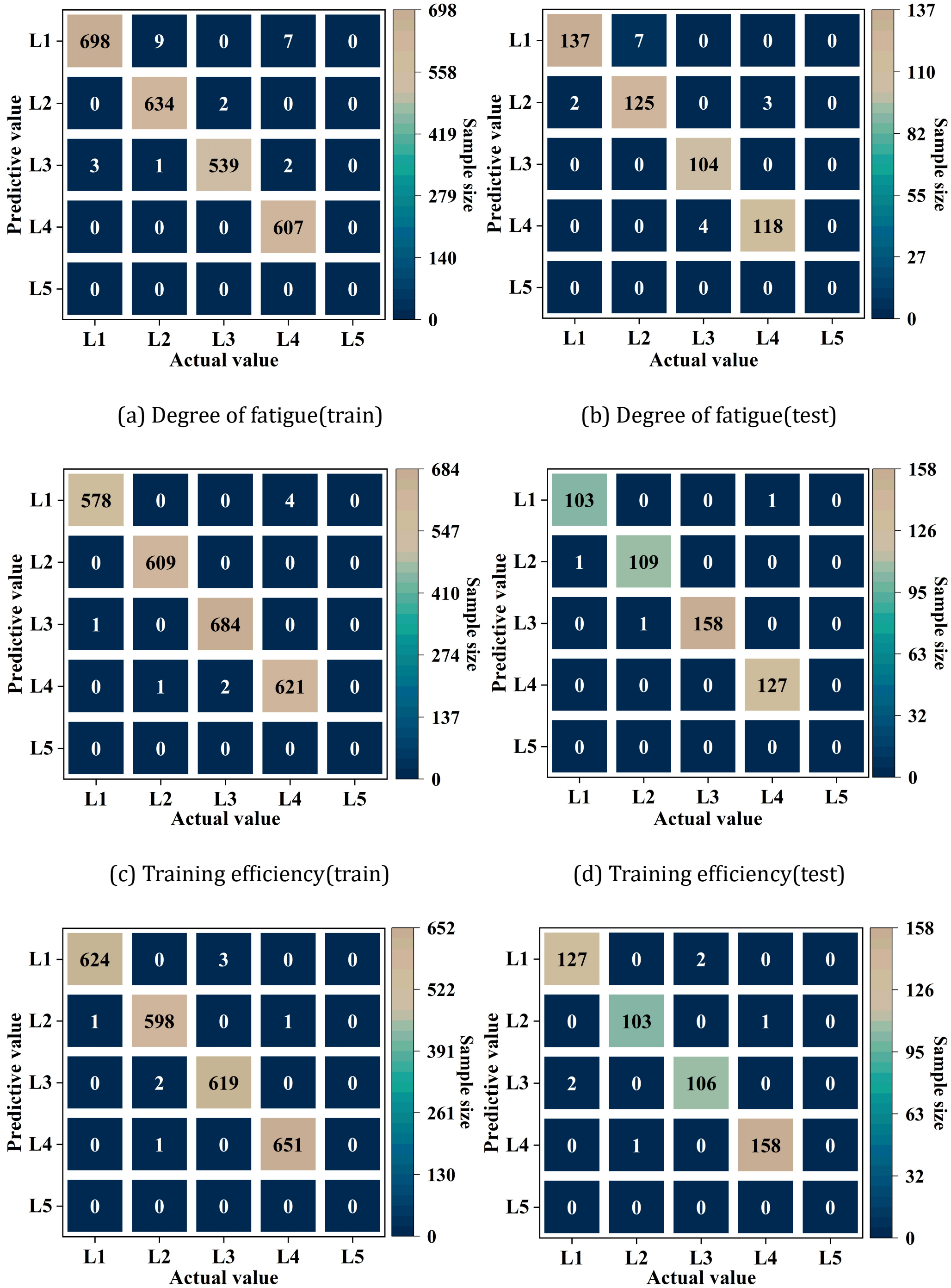
(a) Degree of fatigue (b) Training efficiency

(c) Injury risk (d) Degree of fatigue

(e) Training efficiency (f) Injury risk

Fig. 7. Histogram and scatter plot of absolute error on the entire dataset

After the model is trained, the discrepancy between predicted and actual records must be evaluated. In order to measure the performance of the data model in this paper, the introduction of a confusion matrix is considered, and the results are shown in Fig. 8, with (a)-(f) denoting the results achieved by the three metrics on the training set and the test set, respectively. It is obvious from the figure that the data model constructed based on LSTM network in this paper has good prediction performance. On the fatigue level metric, the prediction accuracy of the model for the training and test sets is 99.12% and 96.80%, respectively. In the model training, 9 L2 samples and 7 L4 samples were predicted as L1 samples, while in the model testing, the same 7 L2 samples were predicted as L1 samples. This indicates that there is still room for optimization of the model for the prediction of the fatigue level, which only has a slight degree of difference. In terms of training efficiency, the LSTM model performs more prominently, with prediction accuracies of 99.68% and 99.40% on the training and test sets, respectively. In the training set, only four L4 samples have more obvious prediction errors, and the rest of the samples as well as the samples on the test set have only a small number of prediction errors. It shows the excellent performance of LSTM model in training efficiency prediction. As for the injury risk metrics, the model's prediction accuracies were 99.68% and 98.80%, respectively. The performance of the model on the test set is slightly inferior compared to the training efficiency index, but the accuracy on the training set can also show the good prediction performance of the model in the injury risk index of youth sports training status. Comprehensive analysis of the above, this paper is based on the use of Internet of Things technology to collect data to build LSTM data model, the model can effectively realize the prediction of youth sports training status, to lay a data foundation for the identification of youth sports training patterns.



(a) Degree of fatigue(train) (b) Degree of fatigue(test)

(c) Training efficiency(train) (d) Training efficiency(test)

(e) Injury risk(train) (f) Injury risk(test)

Fig. 8. The confusion matrix of the prediction results

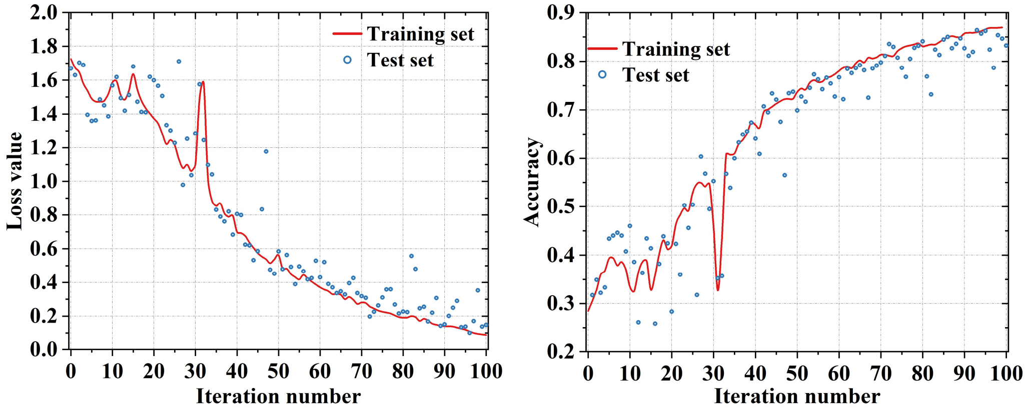
5.2. Pattern Recognition Analysis

5.2.1. Data set segmentation. The data in this section of the experiment still use the 14 adolescents' athletic training dataset obtained in the previous section of the experiment, but the dataset division is slightly different from that in the previous section. In this paper, the athletic training modes of adolescents are categorized into 6 types of sensitive coordination mode, strength development mode, cardiorespiratory endurance mode, balance and stability mode, flexibility and stretching mode, and synergistic training mode, which are denoted as M1-M6, respectively. Therefore, it is necessary to further divide the sample data of the 6 athletic training modes on the basis of 2000 samples in the training set and 500 samples in the test set. The experiment uses the Numerical Python library function in Python to divide the data, and the specific division results are shown in Table 1.

Table 1. Data set

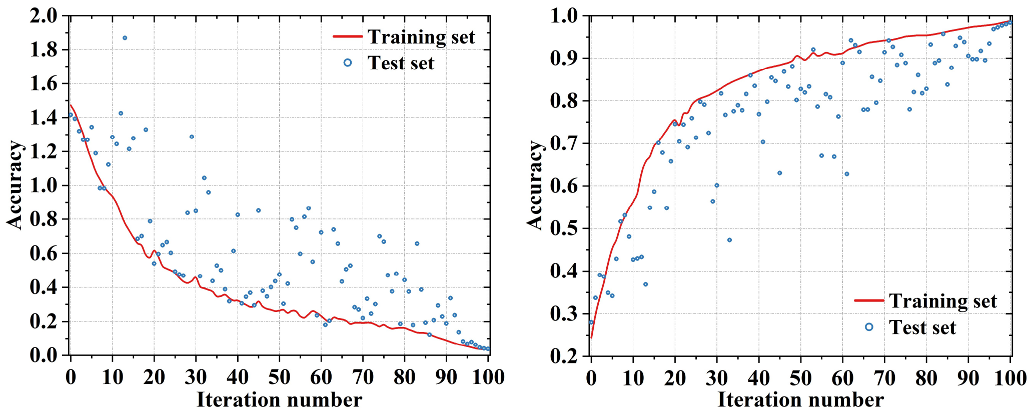
Model	Training set	Test set	Total
M1	347	91	438
M2	339	82	421
M3	341	84	425
M4	312	87	399
M5	354	83	437
M6	307	73	380
Total	2000	500	2500

5.2.2. Experimental results and analysis. In order to illustrate the effectiveness of the youth sports training pattern recognition method constructed based on the SVM algorithm in this paper, a comparative analysis is used to compare the LSTM network with the SVM algorithm in this paper. Firstly, the dataset is input into the LSTM network to get the loss function and recognition accuracy for six patterns, and the results are shown in Fig. 9, (a) and (b) are the loss curve and accuracy curve of the LSTM model, respectively. The final Loss function value of the model on the test set is 0.15 and the accuracy is 83.23%. The dataset was then fed into the SVM model to obtain the loss function and classification accuracy for the six motion training patterns as shown in Fig. 10. After using the same dataset, the same training batch, and the same number of iterations, the final classification accuracy and Loss function were greatly improved, where the Loss function value was reduced to 0.04 and the accuracy on the test set was improved to 98.39%.



(a) Loss (b) Accuracy

Fig. 9. LSTM's identification results

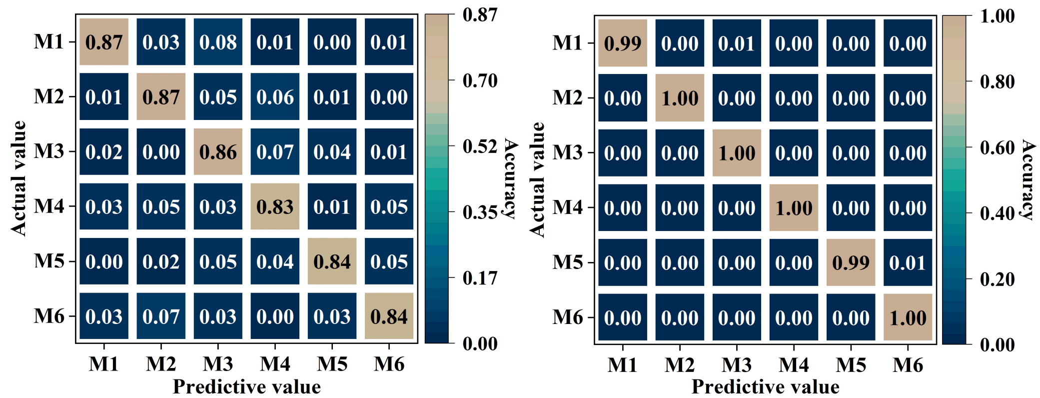


(a) Loss (b) Accuracy

Fig. 10. SVM's identification results

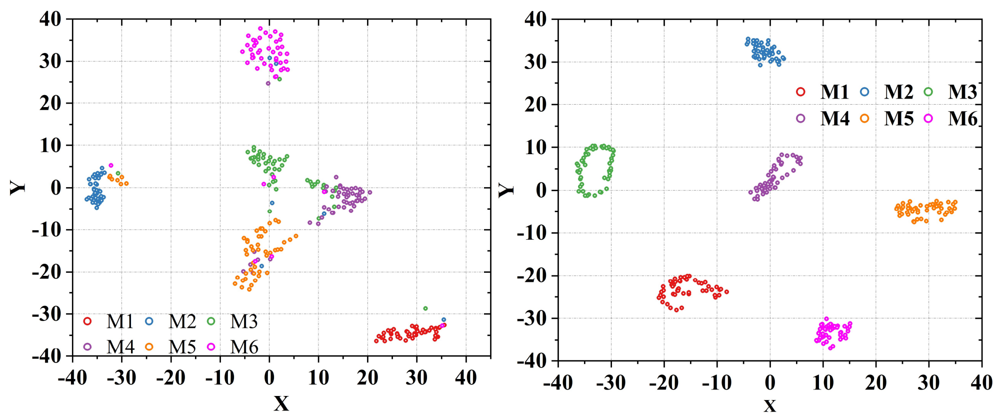
In order to better compare the difference in pattern recognition accuracy between the LSTM model and the SVM in this paper, the recognition results of the trained model on the test set are visualized using the confusion matrix and t-SEN clustering analysis techniques. In this case, the confusion matrix places the predicted results and the true results of all categories by category in the same table, where the number of correct and incorrect recognitions for each category is available; The clustering analysis of the data can better observe the experimental results, find out the relationship between the various categories, and make the data concise, and the t-SNE clustering analysis technology can convert the high-dimensional data into two dimensions, so as to visually judge the performance of the current model. This resulted in the confusion matrix of Fig. 11 and the cluster analysis plot of Fig. 12, where both (a) and (b) show the recognition results of the LSTM and SVM models, respectively. It can be found that the larger error of LSTM in recognizing youth sports training patterns occurs in pattern 4, and 8% of pattern 1 is recognized as pattern 3 in the confusion matrix. Also 6% of mode 2 was identified as mode 4, 7% of mode 3 was identified as mode 4, and 5% of mode 2 was identified as mode 3. In contrast, when using the SVM model to recognize youth sports training patterns, the recognition accuracy of all patterns was improved, and pattern 4, which had the largest error, reached 100% accuracy.

The clustering results likewise demonstrate the ability to obtain more forward pattern recognition performance using the SVM model. The recognition results using the LSTM model in the figure are unsatisfactory with tight distances between the patterns, loose within the patterns, and a significant portion of the patterns are also mixed together. After using the SVM model, the clustering boundaries are obvious, the recognition error is lower, the distance between the patterns is larger, and the interior is tighter. Therefore the clustering effect using SVM model is significantly better than using LSTM model. It shows that the youth sports training pattern recognition model constructed based on SVM algorithm in this paper can adequately recognize the training model based on the collected data as well as the youth sports training state characteristics.



(a) LSTM (b) SVM

Fig. 11. Model recognition accuracy



(a) LSTM (b) SVM

Fig. 12. Model clustering results

6. Conclusion

In this study, we successfully constructed a youth sports training data model in the IoT environment and developed a sports training pattern recognition model based on the support vector machine algorithm. Through the fusion and feature extraction of youth sports training data collected by sensors, the data model was built using LSTM to achieve effective modeling and prediction of youth sports training status. Meanwhile, other mode recognition algorithms were compared and analyzed, and the proposed mode recognition model based on SVM algorithm showed high accuracy and reliability in youth sports training mode recognition. The experimental results show that the models and algorithms proposed in this study can provide scientific data analysis and decision-making support for youth sports training, and have important application value. However, it is found that the data model in this paper has relatively large errors in predicting the “fatigue level” index of youth sports training. Therefore, future research needs to further optimize the data model constructed in this paper to improve the efficiency and accuracy of the algorithm.

Funding

Funded by the Natural Science Foundation of Hainan Province (Project No.: 320MS013).

References

[1] Cerruti, R., Spensieri, V., Presaghi, F., Valastro, C., Fontana, A., & Guidetti, V. (2017). An exploratory study on Internet addiction, somatic symptoms and emotional and behavioral functioning in school-aged adolescents. Clinical Neuropsychiatry, 14(6).

[2] Alagoz, N., & Keskïnkilic, A. U. (2022). The relationship between internet and game addiction and the levels of physical activity of the secondary education students. Medicine Science, 11(1).

[3] Nakshine, V. S., Thute, P., Khatib, M. N., & Sarkar, B. (2022). Increased screen time as a cause of declining physical, psychological health, and sleep patterns: a literary review. Cureus, 14(10).

[4] Shimoga, S. V., Erlyana, E., & Rebello, V. (2019). Associations of social media use with physical activity and sleep adequacy among adolescents: Cross-sectional survey. Journal of medical Internet research, 21(6), e14290.

[5] Xu, T., Zuo, F., & Zheng, K. (2024). Parental Educational Expectations, Academic Pressure, and Adolescent Mental Health: An Empirical Study Based on CEPS Survey Data. International Journal of Mental Health Promotion, 26(2).

[6] Rentala, S., Thimmajja, S. G., Tilekar, S. D., Nayak, R. B., & Aladakatti, R. (2019). Impact of holistic stress management program on academic stress and well-being of Indian adolescent girls: A randomized controlled trial. Journal of education and health promotion, 8(1), 253.

[7] Mohamed, R. A., Mohamad, S. M., & Yusof, H. A. (2024). Impact of exercise on the adolescent immune system: A scoping review. Malaysian Journal of Movement, Health & Exercise, 13(2), 43-56.

[8] Agostinete, R. R., Werneck, A. O., Narciso, P. H., Ubago-Guisado, E., Coelho-e-Silva, M. J., Bielemann, R. M., ... & Vlachopoulos, D. (2024). Resistance training presents beneficial effects on bone development of adolescents engaged in swimming but not in impact sports: ABCD Growth Study. BMC pediatrics, 24(1), 247.

[9] Romadhoni, S., & Yudhistira, D. (2024). Implementing Complex Training Method: Its Effects on Endurance, Speed, Power, and Agility of Adolescent Basketball Players. Physical Education Theory and Methodology, 24(3), 426-432.

[10] Souilla, L., Werner, O., Huguet, H., Gavotto, A., Vincenti, M., Pasquie, J. L., ... & Selly, J. B. (2025). Cardiopulmonary Fitness and Physical Activity Among Children and Adolescents With Inherited Cardiac Disease. JAMA Network Open, 8(2), e2461795-e2461795.

[11] Frömel, K., Šafář, M., Jakubec, L., Groffik, D., & Žatka, R. (2020). Academic stress and physical activity in adolescents. BioMed research international, 2020(1), 4696592.

[12] Beauchamp, M. R., Puterman, E., & Lubans, D. R. (2018). Physical inactivity and mental health in late adolescence. JAMA psychiatry, 75(6), 543-544.

[13] Koval, A. (2017). Ptoblematics of Involvement of Youth in Mass and Professional Sports Activities. Physical education, sport and health culture in modern society, (4 (40)), 10-14.

[14] Schlechter, C. R., Rosenkranz, R. R., Milliken, G. A., & Dzewaltowski, D. A. (2017). Physical activity levels during youth sport practice: does coach training or experience have an influence?. Journal of sports

sciences, 35(1), 22-28.

[15] Post, E. G., Trigsted, S. M., Riekena, J. W., Hetzel, S., McGuine, T. A., Brooks, M. A., & Bell, D. R. (2017). The association of sport specialization and training volume with injury history in youth athletes. The American journal of sports medicine, 45(6), 1405-1412.

[16] Scantlebury, S., Till, K., Sawczuk, T., Phibbs, P., & Jones, B. (2020). Navigating the complex pathway of youth athletic development: Challenges and solutions to managing the training load of youth team sport athletes. Strength & Conditioning Journal, 42(6), 100-108.

[17] Bailón-Cerezo, J., Clarsen, B., Sánchez-Sánchez, B., & Torres-Lacomba, M. (2020). Cross-cultural adaptation and validation of the Oslo Sports Trauma Research Center questionnaires on overuse injury and health problems (2nd version) in Spanish youth sports. Orthopaedic journal of sports medicine, 8(12), 2325967120968552.

[18] Lipman, A., McConnell, E., Kaufman, K., & Giardino, A. (2021). Preventing athlete harm in youth sports. Preventing child abuse: Critical roles and multiple perspectives. Nova Science Publishers.

[19] Weiss, M. R. (2019). Youth sport motivation and participation: Paradigms, perspectives, and practicalities. Kinesiology Review, 8(3), 162-170.

[20] Pandya, N. K. (2021). Disparities in youth sports and barriers to participation. Current reviews in musculoskeletal medicine, 1-6.

[21] Wang, H., & Zhao, B. (2021). Constructing Sports Multi‐Index Data Analysis Based on 5G IoT Technology. Mathematical Problems in Engineering, 2021(1), 6982366.

[22] Castro, R., Mujica, G., & Portilla, J. (2022). Internet of things in sport training: Application of a rowing propulsion monitoring system. IEEE Internet of Things Journal, 9(19), 18880-18897.

[23] Ju, F., Wang, Y., Yin, B., Zhao, M., Zhang, Y., Gong, Y., & Jiao, C. (2023). Microfluidic wearable devices for sports applications. Micromachines, 14(9), 1792.

[24] Zhao, J., Yang, Y., Bo, L., Qi, J., & Zhu, Y. (2024). Research Progress on Applying Intelligent Sensors in Sports Science. Sensors, 24(22), 7338.

[25] Dan, J., Zheng, Y., & Hu, J. (2022). Research on sports training model based on intelligent data aggregation processing in internet of things. Cluster Computing, 25(1), 727-734.

[26] Zhang, Y., Duan, W., Villanueva, L. E., & Chen, S. (2023). Transforming sports training through the integration of internet technology and artificial intelligence. Soft Computing, 27(20), 15409-15423.

[27] Zhao, J., Zhao, Y., & Wang, H. (2023). Simulation of sports training recognition system based on internet of things video behavior analysis. Preventive Medicine, 173, 107611.

[28] Zhen, H., Kumar, P. M., & Samuel, R. D. J. (2021). Internet of things framework in athletics physical teaching system and health monitoring. International Journal on Artificial Intelligence Tools, 30(06n08), 2140016.

[29] Wilkerson, G. B., Gupta, A., & Colston, M. A. (2018). Mitigating sports injury risks using internet of things and analytics approaches. Risk analysis, 38(7), 1348-1360.

[30] Zhao, Y., & You, Y. (2021). Design and data analysis of wearable sports posture measurement system based on Internet of Things. Alexandria Engineering Journal, 60(1), 691-701.

[31] Ping Chen & Jiangui Peng. (2025). Analysis of the sports action recognition model based on the LSTM recurrent neural network. Nonlinear Engineering,14(1).

[32] Lie Yu,Gaotong Hu,Lei Ding,Na Luo & Yong Zhang. (2024). Gait Pattern Recognition based on Multi-sensors Information Fusion through PSO-SVM Model. Engineering Letters,32(5).

[33] Tang X. F. (2024). Integrating empirical mode decomposition and convolutional neural network for

efficient fault diagnosis in metallurgical machinery. Metalurgija,63(3-4),350-352.