Real-Time Collection Method of Athletes’ Abnormal Training Data Based on Machine Learning

Yue Wang

Scientific Research Department, Gansu Police Vocational College, Lanzhou, Gansu 730046, China

Correspondence should be addressed to Yue Wang; wangyue2020@nwnu.edu.cn

Received 5 March 2021; Revised 23 May 2021; Accepted 25 May 2021; Published 3 June 2021

Copyright © 2021 Yue Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Real-time collection of athletes’ abnormal training data can improve the training effect of athletes. This paper studies the real-time collection method of athletes’ abnormal training data based on machine learning. The main motivation of this paper is to collect the athletes’ abnormal training data in time, which can help to evaluate and improve the training effect. Four sensor nodes are arranged in the upper and lower limbs of athletes to collect the angular velocity, acceleration, and magnetic field strength data of athletes in training state. The data are sent to the data transmission base station through wireless sensors, and the data transmission base station transmits the data to the data processing terminal. The data processing terminal calculates the difference between the sample values of each sensor to obtain the data dispersion of each sensor. The features of each dimension data in a time domain and frequency domain are obtained by using the dispersion degree to construct 32-dimensional feature vectors, and the extracted feature vectors are input into the hidden Markov model. The forward algorithm is used to obtain the probability of the final observation sequence, so as to realize the final collection of athletes’ abnormal training data. The experimental results show that the accuracy and recall rate of the abnormal data collected by this method is higher than 98%, which requires less time.

1. Introduction

In competitive sports, the ultimate goal of sports training is to create excellent sports performance. The daily training of athletes is the most basic and controllable factor to improve competitive ability. Athletes’ daily training is the basic way for coaches to understand athletes’ sports conditions. Coaches need to analyze the athletes’ sports situation [1], make clear the training situation of each athlete, evaluate the athletes’ training status according to their own experience, and formulate corresponding training programs to further guide the training and improve the athletes’ sports performance. With the accumulation of athletes’ training data, it has become more and more difficult to manage and analyze these data by manual processing. Using the traditional data processing and database management function of a computer, it can solve the problem of athletes’ training management, help coaches to manage athletes, convert sports performance, manage historical data, and improve the efficiency of data processing [2, 3]. The traditional data analysis and processing methods only analyze the local or surface characteristics of the data and cannot get the description of the overall characteristics of the data which are hidden behind the data and the prediction of its development trend. The collection of athletes’ abnormal training data focuses on the important information hidden behind the data. Data mining technology can extract valuable, unknown, hidden, and potentially useful knowledge from a large number of original data.

In the process of athletes’ training and competition, coaches need to make corresponding training plans according to different athletes’ individual conditions in order to improve athletes’ sports levels. The traditional training method is that coaches make training plans according to their own training theory and experience, combined with the skill level of athletes. This training mode is highly subjective [4]. Coaches need to spend a lot of time analyzing athletes’ posture, and it is difficult to objectively evaluate the training effect of athletes. The core of modern sports training is accuracy and efficiency. If the coach can
accurately control the abnormal data of training, it can greatly improve the effect of sports training. It is a new research direction to collect and analyze athletes’ training data and determine athletes’ abnormal training data, which is of great significance for improving the scientificity of coaches’ training plans and improving athletes’ training effects.

Artificial intelligence is a new comprehensive subject that is developed from computer science, cybernetics, information theory, and other disciplines; it is a science about understanding the internal mechanism of human intelligence and realizing it on the machine. Machine learning is the core content of artificial intelligence research [5–7]. It has been applied in all branches of artificial intelligence, such as natural language understanding, pattern recognition, computer vision, intelligent robot, and other fields [8]. As early as the 1950s, machine learning-related research began, mainly focusing on the connectionist learning of neural networks. From the 1950s to the 1970s, artificial intelligence research was in the “reasoning period,” but with the development of research, it shows that machines with only logical reasoning ability cannot reach artificial intelligence [9]. In the 1980s, machine learning became an independent discipline and began to develop rapidly. Michalski et al. divided machine learning research into “learning in problem solving and planning” and “learning from instructions,” and so on. Feigenbaum divided machine learning technology into four categories in his famous Manual of Artificial Intelligence, namely, “mechanical learning,” “teaching-learning,” “analogue learning,” and “inductive learning”. In the 21st century, machine learning has been applied in various fields. iFLYTEK’s powerful real-time speech recognition technology and today’s headline intelligent news recommendation system are all products of the rapid development of machine learning.

There are many literatures of the evaluation method for athletes based on benefit evaluation theory and regression analysis method which is proposed in reference [10]. The benefit evaluation theory and regression analysis method are used to evaluate the process of athletes’ safety assessment, and the fusion analysis method is used to realize the monitoring and evaluation of physiological indicators. However, this method has poor adaptability to the safety assessment of athlete training and has a large time expenditure. In reference [11], an athlete training safety evaluation model based on big data fusion feature analysis is proposed. The model of integrated information statistics of athlete training safety is constructed and the method of fuzzy association rule scheduling is used to evaluate the safety of athlete training. However, the method carries out the evaluation of the safety of athlete training with a large amount of calculation and the anti-interference is not good. In reference [12], the evaluation method of athlete training safety based on rough set evaluation is proposed.

For the action recognition of human daily behavior, inertial sensors such as gyroscope and accelerometer are mainly used for algorithm classification and pattern recognition of daily behavior such as standing, walking, running, and lying [13]. Wang et al. identified daily activities such as going up and down stairs, running, sitting down, and standing by acquiring data information of five inertial sensors worn in different positions of the body; Wu et al. used deep convolution neural network to classify and identify five kinds of human actions, including walking, sitting down, lying down, running, and standing, and achieved a recognition rate of 0.9126 on Actitraker open source database; Gu et al. used a variety of intelligent sensors in the room to identify the complex daily human behavior. The improved algorithm has a high recognition rate; Atalla used wearable sensors to identify the daily behaviors with different complexity and explored the accuracy of different sensor installation positions under different complexity actions through many experiments. Some researchers tried to use sensors for sports data monitoring and technical evaluation to achieve the effect of auxiliary training. The sensor of human body was used to detect the behavior characteristics of athletes, such as body posture, movement range, and speed, and based on the analysis and mining of athletes’ behavior data, the technical loopholes were found out to help athletes to improve their technical level; Qaisar et al. used multiple acceleration sensors and gyroscopes to identify a variety of different bowling movements, analyze the technical level of action quality by qualitative and quantitative analysis, and make technical evaluation and feedback of bowling posture in bowling training teaching; King et al. designed golf clubs with embedded acceleration sensors. Through receiving data and calculating important parameters related to swing, such as golf club top position, speed, and direction, in the golf training of athletes or amateurs, the data can be analyzed to feed back the quality of users’ swing, so as to achieve the effect of intelligent training.

The basic idea of inertial sensor recognition is that athletes wear simple and light data collection sensors and send the collected data to the processing terminal [14] in real time to identify the athletes’ posture according to various posture data. This method can make up for the lack of image collection and recognition, has low requirements for the use environment and high recognition efficiency, and has become a hot research method of motion attitude recognition. Abnormal data’s real-time collection is an important branch of pattern recognition, which has been widely concerned and developed in recent years. With the rapid development of microelectronics technology, the use of inertial sensors to identify human posture has become a research hotspot. Many researchers apply wearable devices to human auto disturbance recognition. Sensors are used to collect human acceleration, angular velocity, body temperature, heart rate, and much other information. Using the collected information to extract time domain space and frequency space characteristics of athletes’ training actions is convenient to analyze athletes’ abnormal training data [15]. The feature extraction can analyze the athlete’s unit action and transfer the relevant attribute features as sample data to the machine classifier to realize the abnormal data division.

The contributions of this paper are summarized as follows:

This paper studies the real-time collection method of athletes’ abnormal training data based on machine learning.
The sensor is used to collect athlete training data, and the features used for real-time collection of athletes’ abnormal training data are extracted from time domain and frequency domain, respectively. The extracted features are used to accurately collect athletes’ abnormal training data using the hidden Markov model in machine learning. The experimental results verify the effectiveness of this method in real-time collection of athletes’ abnormal training data.

This paper is organized as follows. Section 2 presents the materials and methods. In Section 3, experimental results are presented and analyzed. Finally, Section 4 sums up some conclusions and gives some suggestions as the future research topics.

2. Materials and Methods

2.1. Collection of Sensor Signal. It is the basic condition to collect the abnormal data of athletes’ training accurately to collect the data of human movement posture. In this system, there are four parts, which are sensors, transceiver, processor, and the power supply. The inertial sensor is used to collect human motion posture data. Through the magnetic sensor, angular velocity sensor, and acceleration sensor fixed on the athlete’s body, the data related to human movement is collected [16], and the collected data are transmitted to the terminal processing device for posture recognition through a wireless sensor network. The power supply can provide power to the system. Data quality is the key to affect the accuracy of abnormal data collection in athletes’ training.

The hardware structure of collecting athletes’ abnormal training data is shown in Figure 1.

The hardware structure mainly contains data collection and data transmission, including four data collection nodes and one data transmission base station. The data collection node is composed of three-axis gyroscope MPU3050M, three-axis accelerometer, and magnetometer LSM303DLH, which collect the angular velocity and acceleration data of human body, respectively. The core component of the data sending base station is the wireless transceiver nRF24L01. The receiving node collects data and sends it to the data terminal through the wireless network. The core processing function of data collection module is completed by 32-bit ARM microcontroller STM32F103. The energy supply of the data collection module is provided by a 3.7 V lithium-ion battery.

The data collection signal transmission includes two parts: one is that the sensor node sends the collected human posture data to the data transmission base station; the other is that the data transmission base station sends the data to the processing terminal. The signal transmission between the sensor node and data transmission base station is based on a wireless sensor network. The problem to be overcome is to reduce the data collision rate as far as possible [17], reduce the data loss, and improve the accuracy of data collection. The signal transmission between the data transmission base station and the processing terminal is based on the star topology network, using the time-division multiplexing protocol [18]. It is necessary to calibrate the clock deviation between different nodes to keep the time uniform.

2.2. Feature Extraction of Athletes’ Training Data. After collecting the data of human motion posture, the training data of athletes are divided firstly, and the features of training data are extracted by using the divided data. The extracted features of training data of athletes are sent to the machine learning classifier to realize the collection of abnormal training data of athletes.

2.2.1. Division of Athletes’ Training Data. The degree of dispersion is the difference between the values of the observed variables, and the difference between the sample values of the sensor signal is defined as the degree of dispersion. Taking the angular velocity as an example, \( \omega_n \) represents the \( x \)-axis angular velocity data at the time \( n \), \( \omega_{n-1} \) represents the \( x \)-axis angular velocity data at the time \( n-1 \), and \( d_n \) represents the angular velocity difference between the \( x \)-axis angular velocity of the sensor at the time \( n \) and the previous time. The formula of dispersion \( d_n \) can be obtained as follows:

\[
\text{Dispersion} = \sqrt{\sum (x_i - \mu)^2 / n}
\]

where \( x_i \) represents a single observation value, \( \mu \) represents the mean of the sample data, and \( n \) represents the sample size.
The movement data include angular velocity data and acceleration data [19]. In order to realize the accurate division of athletes’ training data, it is necessary to comprehensively consider the characteristics of each sensor data [20].

\begin{equation}
D_n^a = \sum_{n=0}^{N} a_n^2 + a_n^2 + a_n^2,
\end{equation}

\begin{equation}
D_n^g = \sum_{n=0}^{N} g_n^2 + g_n^2 + g_n^2.
\end{equation}

The movement data include angular velocity data and acceleration data [19]. In order to realize the accurate division of athletes’ training data, it is necessary to comprehensively consider the characteristics of each sensor data [20].

\begin{equation}
D_n^a = \sum_{n=0}^{N} a_n^2 + a_n^2 + a_n^2,
\end{equation}

\begin{equation}
D_n^g = \sum_{n=0}^{N} g_n^2 + g_n^2 + g_n^2.
\end{equation}

The data dispersion of each sensor is calculated, and the athletes’ movement states can be divided by the threshold.

2.2.2. Extraction of Training Data. After data division, the unit action data composed of acceleration and angular velocity are obtained. Acceleration vector sum and angular velocity vector sum are represented by \(a_n\) and \(g_n\), respectively. The formula is as follows:

\begin{equation}
a_n = \sqrt{(a_n^x)^2 + (a_n^y)^2 + (a_n^z)^2},
\end{equation}

\begin{equation}
g_n = \sqrt{(g_n^x)^2 + (g_n^y)^2 + (g_n^z)^2}.
\end{equation}

The three-axis acceleration, three-axis angular velocity, combined acceleration, and combined angular velocity form an 8-dimensional vector, and \(N\) is used to represent the number of sampling points in each unit action, so there are \(N\) sampling data in each dimension of the vector. If each unit action is taken as a sample, then each sample is an \(N \times 8\)-dimensional matrix. The data features of each dimension of each sample are calculated [22], and the extracted signal features include time domain features and frequency domain features. Time domain features include mean value and variance. \(\mu_n\) and \(\delta^2\) are used to represent the mean value and variance of some component of the increment speed of unit action, respectively, and the formula can be obtained as follows:

\begin{equation}
\mu_n = E(a) = \frac{1}{N} \sum_{i=1}^{N} a_i,
\end{equation}

\begin{equation}
\delta^2 = \frac{1}{N} \sum_{i=1}^{N} (a_i - \mu_n)^2,
\end{equation}

where \(a\) is a component of the acceleration.

The frequency domain features include the peak value of discrete Fourier transform and its corresponding frequency [23]. The discrete Fourier transform method is used to transform the signal from time domain to frequency domain. The Fourier transform result of the \(n\)-th sampling point is represented by \(S_{\text{DFT}}(n)\), and the imaginary number unit is represented by \(j\). The formula is as follows:

\begin{equation}
S_{\text{DFT}}(n) = \sum_{i=0}^{N-1} a_i e^{-j(2\pi/N)i}.
\end{equation}

According to the results of Fourier transform, the peak value \(S_{\text{DFT}}(K)\) is obtained. If the sampling point corresponding to the peak value of Fourier transform is \(K\), the corresponding frequency \(f\) formula of Fourier transform is as follows:

\begin{equation}
f = K \times \frac{f_s}{N},
\end{equation}

where \(f_s\) is the sampling frequency of the sensor.

The features of each dimension data in time domain and frequency domain are obtained by feature calculation, and a 32-dimensional feature vector is constructed.
The hidden Markov model is a probability model about time series, which describes the process of generating unobservable state random sequence randomly from a hidden Markov chain and then generating an observation random sequence from each state [24]. The hidden Markov model is determined by the initial probability distribution \( \pi \), the state probability distribution \( A \), and the observation probability distribution \( B \).

Let \( Q = \{ q_1, q_2, \ldots, q_N \} \) be the set of all possible states and \( V = \{ v_1, v_2, \ldots, v_M \} \) be the set of all possible observations, where \( N \) is the number of possible states and \( M \) is the number of possible observations; \( I = \{ i_1, i_2, \ldots, i_T \} \) is the state sequence of length \( T \) and \( O = \{ o_1, o_2, \ldots, o_T \} \) is the corresponding observation sequence.

The formula of state transition probability matrix \( A \) is as follows:

\[
A = \begin{bmatrix} a_{ij} \end{bmatrix}_{N \times N},
\]

where \( a_{ij} = P(i_{t+1} = q_j | i_t = q_i), \ i = 1, 2, \ldots, N \) and \( j = 1, 2, \ldots, N \), and \( a_{ij} \) is the probability of transition from state \( q_i \) at time \( t \) to state \( q_j \) at time \( t + 1 \).

The formula of the observed probability matrix \( B \) can be obtained as follows:

\[
B = \begin{bmatrix} b_j(k) \end{bmatrix}_{N \times M},
\]

where \( b_j(k) = P(o_k = v_j | i_t = q_j), \ k = 1, 2, \ldots, M \), \( j = 1, 2, \ldots, N \), and \( b_j(k) \) is the probability of generating observation \( v_k \) when it is in state \( q_j \) at time \( t \). \( \pi \) is the initial state probability vector; \( \pi_i = P(i_1 = q_i) \) is the probability that \( t = 1 \) is in state \( q_i \).

Hidden Markov model \( \lambda \) can be represented by the following symbols, namely,

\[
\lambda = (A, B, \pi).
\]

In formula (10), \( A, B, \) and \( \pi \) are called three elements of the hidden Markov model.

For a given hidden Markov model \( \lambda = (A, B, \pi) \) and observation sequence \( O = \{ o_1, o_2, \ldots, o_T \} \), the probability \( P(O | \lambda) \) of the observation sequence in the model is calculated, and the final probability result is obtained by forward algorithm.

Firstly, forward probability is defined. In the model \( \lambda = (A, B, \pi) \), the probability that part of the observation sequence at time \( t \) is \( o_1, o_2, \ldots, o_t \) and the state is \( q_j \) is defined as the forward probability, which is denoted as

\[
a_t(i) = P(o_1, o_2, \ldots, o_t, i_t = q_i | \lambda).
\]

The process of obtaining the observation sequence probability \( P(O | \lambda) \) is as follows:

1. Initial value

\[
a_1(i) = \pi b_i(o_1).
\]

2. For \( t = 1, 2, \ldots, T - 1 \), the formula is as follows:

\[
a_{t+1}(i) = \sum_{j=1}^{N} a_t(j) a_{ji} b_j(o_{t+1}).
\]

(13)

(3) End

\[
P(O | \lambda) = \sum_{i=1}^{N} a_T(i).
\]

(14)

The forward algorithm is to calculate the probability \( P(O | \lambda) \) of observation sequence through the known hidden Markov model \( \lambda = (A, B, \pi) \) and observation sequence \( O = \{ o_1, o_2, \ldots, o_T \} \). \( a_t(j) \) is the forward probability when the observation is \( o_1, o_2, \ldots, o_t \) and the state is \( q_j \) at the time \( t \), and \( a_t(j) a_{ji} \) is the joint probability when the observation is \( o_1, o_2, \ldots, o_t \) and the state is \( q_j \) at the time \( t \) and \( q_i \) at the time \( t + 1 \). The sum of all possible \( N \) states \( q_j \) at the time \( t \) is the joint probability of \( o_1, o_2, \ldots, o_t \) at the time \( t \) and \( q_j \) at the time \( t + 1 \). The product of this result and the observation probability \( b_j(o_{t+1}) \) is just the forward probability \( a_{t+1}(i) \) when the observation is \( o_1, o_2, \ldots, o_t \) and the state is \( q_j \) at the time \( t + 1 \). For \( t = T \), there are

\[
a_T(i) = P(o_1, o_2, \ldots, o_T, i_T = q_i | \lambda).
\]

(15)

So, the probability of the final observation sequence is \( P(O | \lambda) = \sum_{i=1}^{N} a_T(i) \).

The efficiency of the forward algorithm is to recursively deduce the forward probability to the global by using the path structure of the model to get the final probability \( P(O | \lambda) \). Each recursion directly refers to the calculation result of the previous time [22], which avoids repeated calculation and reduces the time complexity of the algorithm from \( O(TN^T) \) to \( O(N^2T) \).

3. Results

In order to verify the effectiveness of the real-time collection method for abnormal data of athletes’ training, 8 basketball players of sports major in a university are selected as the experimental objects, and 9 training movements including walking, running, jumping, when there is no ball, standing dribbling, walking dribbling, running dribbling, shooting, passing, and catching when holding the ball are made. The sensors are placed on the upper and lower limbs of the athletes. A total of 12000 samples are collected, including 6000 upper limb movements and 6000 lower limb movements of standing dribble, walking dribble, shooting, passing, catching, and running dribble. During the sample collection process, the subjects completed the training according to the regulations.

The specific contents of the samples are shown in Table 1.

Considering that the real-time collection of athletes’ abnormal training data is a binary classification problem, the collection accuracy, collection recall rate, F1 value, mean
Table 1: Athlete training actions and sample size.

| Sports parts | Action                  | Number of samples (N) |
|--------------|-------------------------|-----------------------|
| Upper limb   | Standing dribble        | 1000                  |
|              | Walking dribble         | 1000                  |
|              | Running dribble         | 1000                  |
|              | Shot                    | 1000                  |
|              | Pass                    | 1000                  |
|              | Catch the ball          | 1000                  |
| Lower limbs  | Walk                    | 1000                  |
|              | Run                     | 1000                  |
|              | Jump                    | 1000                  |
|              | Shot                    | 1000                  |
|              | Running dribble         | 1000                  |
|              | Walking dribble         | 1000                  |
| Total        |                         | 12000                 |

square error, and AUC (area under the curve) value are used to measure the real-time collection performance of athletes’ abnormal training data. The accuracy of collection indicates the proportion between the correct collection of abnormal data instances and all instances assigned to the class by the classifier. Recall rate represents the proportion of instances in a given category correctly classified by the machine learning classifier. The $F_1$ value is a harmonic average of precision and recall. AUC is the probability that the positive instance selected randomly by the classifier is higher than the negative instance selected randomly (assuming that “positive” is higher than “negative”). When AUC is close to 1, it means that the collection accuracy is higher. When AUC is close to 0.5, it means that machine learning is a random classification condition, and the collection accuracy is poor. BP neural network and support vector machine are selected as comparison methods, and the above two methods are compared with the method in this paper.

Three methods are used to collect the peak signal-to-noise ratio of athletes' training data sample, the comparison results are shown in Table 2. The BP neural network method used in the experiment is proposed in reference [25]. The support vector machine method used in the experiment is proposed in reference [26].

The experimental results in Table 2 show that the peak signal-to-noise ratio of the sample signals collected by the proposed method is higher than 30 dB; the peak signal-to-noise ratio of the sample signals collected by BP neural network method is less than 30 dB; the peak signal-to-noise ratio of the sample signals collected by support vector machine method is less than 29 dB. The experimental results show that the quality of the action signal collected by the proposed method is significantly higher than that of the other two methods. The high quality of the signal collected by this method helps to accurately extract the training characteristics of athletes and provides a theoretical basis for the accurate collection of abnormal data of athletes’ training.

The comparison results of the accuracy of collecting athletes' abnormal training data by three methods are shown in Figure 3.

The experimental results in Figure 3 show that the accuracy of the proposed method is significantly higher than that of the other two methods. The accuracy of the abnormal data collected by the proposed method is higher than 98.5%; the accuracy of the abnormal data collected by BP neural network method and support vector machine method is lower than 96%. The performance of support vector machine method is the worst. According to Figure 3, the accuracy of abnormal data collection in athlete training of the method proposed in this paper is 4.3% higher than the BP neural network method on average and is 6.5% higher than the support vector machine method. So, we can see that the proposed method has high accuracy of abnormal data collection in athlete training and can be applied to the abnormal data collection in actual athlete training.

The comparison results of recall rate of athletes’ abnormal training data collected by three methods are shown in Figure 4.

As can be seen from the experimental results in Figure 4, the recall rate of abnormal data collected by the proposed method is significantly higher than that of the other two methods. The recall rate of abnormal training data collected by the proposed method is higher than 98%; the recall rate of abnormal training data collected by BP neural network method and support vector machine method is lower than 97%. The performance of support vector machine method is the worst. According to Figure 4, the recall rate of abnormal data collection in athlete training of the method proposed in this paper is 4.7% higher than the BP neural network method on average and is 6.1% higher than the support vector machine method. So, we can see that the proposed method has high recall rate of abnormal data collection and superior collection performance.

The comparison results of $F_1$ value of athletes’ abnormal training data collected by three methods are shown in Table 3.

$F_1$ value is an important index to measure the accuracy and recall rate of abnormal data collection. The closer the $F_1$ value of abnormal data collection is to 1, the better the collection performance is. The experimental results in Table 3 show that the $F_1$ value of athletes’ abnormal training data collected by the proposed method is significantly higher than that collected by the other two methods. The $F_1$ value of abnormal data collected by the proposed method is higher than 0.93, very close to 1; the $F_1$ value of abnormal data collected by BP neural network method and support vector machine method is lower than 0.86. The results show that the proposed has high accuracy, high recall, and high reliability, which can provide a theoretical basis for coaches to make training plans.

The comparison results of mean square error (MSE) of athletes’ abnormal training data collected by three methods are shown in Figure 5.

Experimental results in Figure 5 show that the mean square error of the abnormal data collected by the proposed method is significantly lower than that of the other two methods. The results show that the mean square error of abnormal data collected by the proposed method is lower than 0.04; the mean square error of abnormal data collected by BP neural network method and support vector machine method is higher than 0.05. The results show that the proposed method has low mean square error of abnormal data collection and high reliability.
AUC comparison results of athletes’ abnormal training data collected by three methods are shown in Figure 6.

As can be seen from the experimental results in Figure 6, the AUC value of abnormal data collected by the proposed method is very close to 1, while the AUC value of abnormal data collected by BP neural network method and support vector machine method is very close to 0.5. The results show that the accuracy of the proposed method to collect abnormal data of athletes training is high, and the method of BP neural network and support vector machine to collect abnormal data of athletes training is mostly random classification. The method of this paper to collect abnormal data of athletes training has high accuracy and high reliability.

The abovementioned experimental results effectively verify that the method in this paper has high accuracy in collecting abnormal data of athletes’ training and has high performance in collecting abnormal data of athletes’ training, which can be applied to the practical application of athletes’ training. In order to further verify the real-time ability of collecting abnormal training data of athletes by this method, the proposed method is used to collect athletes’

| Action            | Method of this article (dB) | BP neural network method (dB) | Support vector machine method (dB) |
|-------------------|-----------------------------|------------------------------|-----------------------------------|
| Standing dribble  | 35.61                       | 27.52                        | 25.61                             |
| Walking dribble   | 34.25                       | 26.48                        | 26.34                             |
| Running dribble   | 36.52                       | 25.61                        | 27.85                             |
| Shot              | 34.15                       | 26.85                        | 24.05                             |
| Pass              | 35.62                       | 27.52                        | 25.64                             |
| Catch the ball    | 34.85                       | 28.61                        | 23.61                             |
| Walk              | 36.25                       | 25.64                        | 27.85                             |
| Run               | 33.45                       | 26.34                        | 26.58                             |
| Jump              | 31.52                       | 29.52                        | 28.64                             |
| Shot              | 34.52                       | 24.81                        | 26.34                             |
| Running dribble   | 33.65                       | 23.52                        | 21.52                             |
| Walking dribble   | 34.85                       | 26.48                        | 23.96                             |

![Figure 3: Comparison of collection accuracy.](image-url)
abnormal training data. The comparison results of collection time under different data samples are shown in Table 4.

Table 4 shows that the time cost of using the proposed method to collect athletes’ abnormal training data is the lowest under different sample numbers. The comparison results show that the proposed method can quickly collect athletes’ abnormal training data in a short time. This method not only has high accuracy but also needs less time to collect abnormal data. It can quickly obtain accurate abnormal data and has high practicability.

In this paper, we propose a real-time collection method of athletes’ abnormal training data based on machine learning. According to [27], this paper proposed a HMM-based asynchronous $H^\infty$ filtering for fuzzy singular Markovian switching systems with retarded time-varying delays. The computational complexity of the method proposed by our paper mainly depends on the characteristics of the training network, while the computational complexity of the method proposed in [27] is mainly dependent on the HMM method. So, we can draw that the performance of the method proposed by us is much better.

Table 3: F1 value comparison.

| Training action     | Method of this article | BP neural network method | Support vector machine method |
|---------------------|------------------------|--------------------------|-------------------------------|
| Walk                | 0.978                  | 0.758                    | 0.658                         |
| Run                 | 0.986                  | 0.851                    | 0.711                         |
| Jump                | 0.984                  | 0.762                    | 0.658                         |
| Standing dribble    | 0.986                  | 0.714                    | 0.725                         |
| Walking dribble     | 0.934                  | 0.824                    | 0.714                         |
| Running dribble     | 0.946                  | 0.762                    | 0.725                         |
| Shot                | 0.945                  | 0.824                    | 0.685                         |
| Pass                | 0.952                  | 0.841                    | 0.765                         |
| Catch the ball      | 0.967                  | 0.765                    | 0.715                         |
Figure 5: Comparison of mean square error.

Figure 6: AUC comparison.
4. Conclusions

With the development of wireless sensor network and microelectronic equipment technology, athletes’ abnormal training data collection has been widely concerned in various fields. The sensor equipment is used to collect the athletes’ upper and lower limb movement state signals to extract the athletes’ training characteristics, and the hidden Markov model is used in machine learning to complete the effective collection of athletes’ abnormal training data. Selecting basketball players as the experimental object, the abnormal data collection of athletes’ training is realized in the field of basketball. The experimental results effectively verify that the method is highly effective in collecting abnormal data of athletes’ training. The research results provide a new collection scheme for abnormal data of sports training. The collected dataset will be made publicly available by the other researchers. For this new system, when we use it, it has a big structure size, which is not very easy to carry, so in order to facilitate large-scale use, its size must be reduced. In addition, we can speed up data processing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Q. L. Liu and X. P. Wang, “Real time monitoring simulation of abnormal state of large data in smart grid,” Computer Simulation, vol. 36, no. 3, pp. 364–367, 2019.
[2] H. Yao, P. Gao, P. Zhang, J. Wang, C. Jiang, and L. Lu, “Hybrid intrusion detection system for edge-based IoT relying on machine-learning-aided detection,” IEEE Network, vol. 33, no. 5, pp. 75–81, 2019.
[3] Z. Wu, D. Rincon, and P. D. Christofides, “Real-time adaptive machine-learning-based predictive control of nonlinear processes,” Industrial & Engineering Chemistry Research, vol. 59, no. 6, pp. 2275–2290, 2020.
[4] H. A. Yavasoglu, Y. E. Tetik, and K. Gokce, “Implementation of machine learning based real time range estimation method without destination knowledge for bevs,” Energy, vol. 172, pp. 1179–1186, 2019.
[5] M. Yang, B. Ai, R. He, C. Huang, and J. Li, “Machine-learning-based fast angle-of-arrival recognition for vehicular communications,” IEEE Transactions on Vehicular Technology, vol. 70, no. 2, pp. 1592–1605, 2021.
[6] Y. Xu, D. Li, Z. Wang, Q. Guo, and W. Xiang, “A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals,” Wireless Networks, vol. 25, no. 7, pp. 3735–3746, 2017.
[7] D. Liu, G. Zhu, Q. Zeng, J. Zhang, and K. Huang, “Wireless data collection for edge learning: data-importance aware retransmission,” in Proceedings of the 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 1–6, Cannes, France, July 2019.
[8] X.-Q. Wang and J. Yin, “Application of machine learning in safety evaluation of athletes training based on physiological index monitoring,” Safety Science, vol. 120, pp. 833–837, 2019.
[9] G. Verrall, A. Hains, B. Ayres, and R. Hillock, “Influence of type and duration of training on the presence of an abnormal ECG in high-performance athletes,” Heart Asia, vol. 11, no. 1, Article ID e011120, 2019.
[10] M. Fakhar, M. R. Mahyarinia, and J. Zafarani, “On non-smooth robust multobjective optimization under generalized convexity with applications to portfolio optimization,” European Journal of Operational Research, vol. 265, no. 1, pp. 39–48, 2018.
[11] G. Farnadi, S. H. Bach, M.-F. Moens, L. Getoor, and M. De Cock, “Soft quantification in statistical relational learning,” Machine Learning, vol. 106, no. 12, pp. 1971–1991, 2017.
[12] A. Rames, M. Rodriguez, and L. Getoor, “Multi-relational influence models for online professional networks,” in Proceedings of the 2017 International Conference on Web Intelligence, ACM, pp. 291–298, Leipzig, Germany, August 2017.
[13] J. F. L. de Oliveira, E. G. Silva, and P. S. G. de Mattos Neto, “A hybrid system based on dynamic selection for time series forecasting,” IEEE Transactions on Neural Networks and Learning Systems, vol. 99, pp. 1–13, 2021.
[14] C. Rea, K. J. Montes, K. G. Erickson, R. S. Granetz, and R. A. Tinguely, “A real-time machine learning-based disruption predictor in DIII-D,” Nuclear Fusion, vol. 59, no. 9, Article ID 096016, 2019.
[15] G. Daniel, F. Ceraudo, O. Limousin, D. Maier, and A. Meuris, “Automatic and real-time identification of radionuclides in gamma-ray spectra: a new method based on convolutional neural network trained with synthetic data set,” IEEE

| Number of samples (N) | Method of this paper (ms) | BP neural network method (ms) | Support vector machine method (ms) |
|-----------------------|---------------------------|------------------------------|----------------------------------|
| 1000                  | 32                        | 105                          | 135                              |
| 2000                  | 56                        | 124                          | 182                              |
| 3000                  | 72                        | 135                          | 234                              |
| 4000                  | 82                        | 164                          | 294                              |
| 5000                  | 91                        | 185                          | 335                              |
| 6000                  | 105                       | 205                          | 394                              |
| 7000                  | 111                       | 234                          | 435                              |
| 8000                  | 135                       | 268                          | 497                              |
| 9000                  | 159                       | 342                          | 581                              |
| 10000                 | 175                       | 394                          | 645                              |
| 11000                 | 235                       | 428                          | 785                              |
| 12000                 | 268                       | 594                          | 856                              |
[16] M. Al-Saud, A. M. Eltamaly, M. A. Mohamed, and A. Kavousi-Fard, “An intelligent data-driven model to secure intravehicle communications based on machine learning,” IEEE Transactions on Industrial Electronics, vol. 67, no. 6, pp. 5112–5119, 2020.

[17] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, “A probability density function generator based on neural networks,” Physica A: Statistical Mechanics and Its Applications, vol. 541, Article ID 123344, 2020.

[18] S. Sadiqbatcha, J. Zhang, H. Zhao, H. Amrouch, and X. D. Tan, “Post-silicon heat-source identification and machine-learning-based thermal modeling using infrared thermal imaging,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 40, no. 4, pp. 694–707, 2020.

[19] D. Xiao, F. Fang, J. Zheng, C. C. Pain, and I. M. Navon, “Machine learning-based rapid response tools for regional air pollution modelling,” Atmospheric Environment, vol. 199, no. 15, pp. 463–473, 2019.

[20] Y. Yang, J. Lu, C. Yang, and Y. Zhang, “Exploring fragment-based target-specific ranking protocol with machine learning on cathepsin S,” Journal of Computer-Aided Molecular Design, vol. 33, no. 12, pp. 1095–1105, 2019.

[21] G. Burlak, “Evaluation of the spectrum of a quantum system using machine learning based on incomplete information about the wavefunctions,” Applied Physics Letters, vol. 116, no. 2, Article ID 024101, 2020.

[22] W. Yuan, G. Han, and D. Guan, “Learning from mislabeled training data through ambiguous learning for in-home health monitoring,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 549–569, 2020.

[23] B. B. Thomas, Y. Mou, L. Keeler, C. Magnan, and S. Agersborg, “A highly sensitive and specific gene fusion algorithm based on multiple fusion callers and an ensemble machine learning approach,” Blood, vol. 136, no. 1, pp. 12-13, 2020.

[24] M. J. Bianco and P. Gerstoft, “Unsupervised machine learning for acoustic tomography and event detection,” The Journal of the Acoustical Society of America, vol. 146, no. 4, Article ID 2845, 2019.

[25] T. Li, J. Sun, and L. Wang, “An intelligent optimization method of motion management system based on BP neural network,” Neural Computing and Applications, vol. 33, no. 4, pp. 707–722, 2020.

[26] X. Liang, T. Qi, T. Qi, Z. Jin, and W. Qian, “Hybrid support vector machine optimization model for inversion of tunnel transient electromagnetic method,” Mathematical Biosciences and Engineering, vol. 17, no. 4, pp. 3998–4017, 2020.

[27] G. Zhuang, W. Sun, S.-F. Su, and J. Xia, “Asynchronous feedback control for delayed fuzzy degenerate jump systems under observer-based event-driven characteristic,” IEEE Transactions on Fuzzy Systems, vol. 1, no. 99, pp. 1–15, 2020.