International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-21

https://doi.org/10.70917/ijcisim-2025-0203
Article

Research on Student Behavior Pattern Recognition in Physical Education Based on Adaptive Algorithm

Xiaolu Li [1], Junyi Yang [2] and Xiaoguang Yang [1,*]

[1] College of Physical Education, Yan'an University, Yan'an, Shaanxi, 716000, China
[2] Northwest A & F University, Yangling, Shaanxi, 712100, China; 137957462@163.com

Abstract: In order to address the need for accurate identification of student behaviors in physical education, this paper conducts research on student behavior pattern recognition based on adaptive algorithms. Firstly, the human skeleton spatio-temporal motion model is used as a tool to transform continuous actions into structured data. Then multi-intelligent body deep reinforcement learning is used to realize the parallel recognition and decision-making of multi-threaded behaviors. On this basis, a multi-scale spatio-temporal feature encoder is built to capture behavioral features at different scales. Finally, the features are fused and optimized with the help of graph convolution adaptive reinforcement network (DT-AGCN) to complete the adaptive reinforcement recognition of students' behavioral patterns. The experimental results show that on the KTH dataset, the model achieves an average recognition accuracy of 98.12% for ten types of basic sports actions, and the recognition accuracy remains above 93.11% even in the complex scene environment with partial occlusion. In the more challenging UCF dataset, the model still achieves 95.84% of the overall recognition accuracy for nine categories of teaching activities in a real classroom with complex scenarios. In the 12-week empirical study, it can be seen that students' sports participation increased from 58.51% to 98.93% at the beginning of the period, the skill attainment rate increased to 97.54%, the physical fitness pass rate increased from 62.83% to 99.03%, and the number of classroom interactions increased to 72, which is a strong proof that the teaching model using the model can greatly enhance the effectiveness of teaching and learning.

Keywords: behavior recognition; adaptive algorithm; physical education; graph convolutional network; multi-scale features

1. Introduction

Physical education is an important part of students' comprehensive quality education, which plays an important role in cultivating students' physical fitness, coordination ability and teamwork spirit. However, due to the differences in age characteristics and physical development, the different behavioral patterns demonstrated by students participating in classroom learning in physical education have a great impact on the teaching effect [1-3]. Therefore, in order to provide students with better learning environments and guidance, as well as to provide basic support for teachers to develop teaching strategies suitable for different students, teachers need to identify and analyze students' learning behavior patterns [4-6].

In order to better analyze students' learning behavior patterns, it is necessary to obtain multiple sources of information from students for a comprehensive analysis. The main sources of information include students' self-reports, teachers' observations, and records of learning achievements, from which the general characteristics of learners' learning behavior patterns are derived [7-8]. However, these methods are task-heavy, inefficient, and inaccurate; and with the development of artificial intelligence, as an effective optimization algorithm, adaptive algorithms can play an important role in the identification of student behavior patterns [9-11]. Adaptive algorithms are algorithms that can be automatically adjusted according to environmental changes and problem situations, and they are widely used in the

fields of computer science and artificial intelligence to solve a variety of complex problems, including optimization problems, pattern recognition, learning systems, etc. [12-15]. The design of these algorithms is often inspired by the adaptive ability of biology, such as genetic algorithms, simulated annealing algorithms, particle swarm algorithms, and so on. Adaptive algorithms are widely used to solve a variety of complex problems, including optimization problems, pattern recognition, learning systems, etc [16-17]. In the recognition of students' behavioral patterns in physical education, adaptive algorithms can realize the personalized adaptation of teaching strategies through real-time data collection and analysis, laying a solid foundation for improving the effectiveness of physical education teaching [18-21].

In the context of the continuous development and application of artificial intelligence, student behavior pattern recognition is not limited to adaptive algorithms, in which neural networks, deep learning, and computer vision are also widely used. Literature [22] examined the application of neural networks in student behavior pattern recognition in teaching and learning, aiming to make choices that can complete the educational environment and bring the best results for students and educators. Literature [23] developed a facial feature point localization method based on deep convolutional neural networks and cascade techniques applied to student behavioral patterns is gender, which was verified to have excellent performance by analyzing and identifying head pose and facial expression. Literature [24] emphasized the importance of classroom behavior data in assessing and predicting students' academic performance, proposed a moving target detection algorithm for identifying student behaviors in the classroom, and experimentally verified the effectiveness of the algorithm in identifying student behaviors in the classroom. Literature [25] analyzed the application of machine learning techniques in the recognition of students' behavioral patterns, aiming to improve the efficiency of teaching supervision and management, and based on experiments it was pointed out that the machine learning model performs well in the recognition task, but it needs to be further improved. Literature [26] proposed a scheme for recognizing students' classroom teaching behaviors based on a dual-stream convolutional neural network model, and constructed an attention mechanism combining the dual-stream convolutional neural network model to improve the accuracy of the model, revealing that the recognition ability performs well under the combination of the two. Literature [27] proposed a deep learning recognition method applicable to abnormal student behaviors in intelligent classroom teaching scenarios, verified the effectiveness of the designed method, and its average false alarm rate and recognition delay time performed well. Literature [28] in order to improve the effectiveness of distance education student state analysis, combined with the human skeletal model real-time recognition of student behavior in the classroom constructed a network topology model, and verified the effectiveness of the human skeletal model based on the human skeletal model. Literature [29] introduced a deep learning based automation framework for identifying classroom activities covering student and teacher behaviors from classroom videos, aiming to automate the identification of classroom behaviors and improve the assessment of educational quality. Literature [30] proposed a new feature fusion network for student behavior recognition and combined a spatial affine transform network with a convolutional neural network to extract more detailed features, revealing that the proposed algorithm performs well for behavior recognition. Literature [31] devised a computer vision based approach to identify student behavior in the classroom and by using it in practice showed that the approach performs excellently in identifying student behavior in the classroom with more than 95% accuracy.

In the article, the initial behavior recognition model is constructed from the nature of human movement, and then the multi-scale feature extraction technique is adopted to enhance the model's ability to recognize complex spatial and temporal features. Finally, the adaptive enhancement mechanism of graph convolutional network is used to merge and optimize the above features to form a set of student behavior pattern recognition system with strong generalization ability in physical education. Specifically, the human skeleton spatio-temporal motion model transforms a sequence of student actions into structured data, and the multi-intelligence deep reinforcement learning algorithm splits the complex task into multiple parallel computational intelligences to achieve efficient identification and preliminary determination of multi-threaded, concurrent behaviors, based on which a proprietary multi-scale spatio-temporal feature encoder is designed to separate spatio-temporal pooling of attention and other initiatives, so that the model can simultaneously focus on local subtle actions and global motion trajectories, and then accurately capture different scales of behavioral features such as shooting shots and overall running position. Finally, we adopt the graph convolution adaptive enhancement network to fuse the features, which is based on the natural graph structure of the human skeleton data, and explicitly establishes the dependency model between the joints with the graph convolution network (GCN). Relying on its internal adaptive enhancement mechanism and the T-DT temporal dimensionality conversion module, the network can dynamically adjust its own parameters to realize the high-precision discrimination of the student's behavioral patterns.

2. Adaptive algorithm based technology for recognizing students' motor behavior

2.1. Human motion behavior recognition algorithm

2.1.1. Modeling the spatio-temporal motion of the human skeleton

In the original image samples, in order to extract the skeletal point information, it is necessary to first extract the feature points and build a spatio-temporal motion model. During the execution of this extraction step, the network can be divided into upper and lower 2 branches, and the optimization of the skeletal features will be completed by feature fusion if there is stage localization in the relevant region. The result of the network is shown in equation (1).

$$\begin{cases} X = \rho^T (f + l) \\ Y = \rho^T (f - l) \end{cases} \tag{1}$$

Where: X and Y denote the X and Y coordinate points in the network model, ρ^T denotes the region confidence matrix, and f and l denote the input features and output features of the association field, respectively.

The network model is built by Eq. (1), the cleaning and frame extraction of skeletal points are completed, and many difficult to accurately detect skeletal points are used as the training samples and redundant frames of the subject, and the one with the longest duration among all the samples is selected, and the zero-completion process is completed in order to better deal with the truncation and directional partitioning work. In the skeletal points, to ensure that all features can retain information, set up the skeletal point model as shown in Fig. 1.

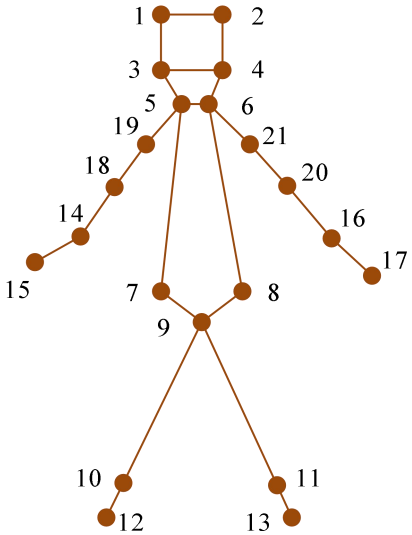


Figure 1. Bone point model.

In Fig. 1, 21 skeletal point coordinates are set in the skeletal point-point model, with point 1 and point 2 indicating the skull, point 3 and point 4 indicating the orbit, point 5 and point 6 indicating the mandible, point 7 and point 8 indicating the hip bone, point 9 indicating the tail bone, point 10 with point 11 indicating the kneecap bone, point 12 with point 13 indicating the ankle bone, point 14 with point 16 indicating the carpal bone, point 15 with point 17 indicating the phalanges, point 18 with point 20 indicate the ulna, and dots 19 and 21 indicate the humerus. Once the human body moves in the image, the points will also move, so the fluctuation of the skeletal points needs to be reduced to ensure the accuracy of the data. At this time, the formula required for the processing of skeletal points is shown in equation (2).

$$x_{move} = A_k h_n \tag{2}$$

Where: x_{move} denotes the length of the skeletal point feature movement, A_k denotes the total number of nodes of the skeletal point, and h_n denotes the length of the vector in each frame.

Combining Eq. (2) with the spatio-temporal vectors, a spatio-temporal motion model of the human skeleton can be built so that the algorithm can obtain the coordinates of the skeletal points in the moving

image.

2.1.2. Multi-threaded recognition results based on multi-intelligence deep reinforcement learning computation

In multi-intelligence deep reinforcement learning algorithms, the intelligences usually have the ability to transfer from one behavior (or action) to another, which is reflected in their ability to select and execute the corresponding action according to the current state of the environment as well as their own strategy, thus realizing the transition between the states, and this transfer information k_e is shown in equation (3).

$$k_e = \{x, y, z, h\} \tag{3}$$

Where: x denotes the state in horizontal coordinates, y denotes the state in vertical coordinates, z denotes the state in spatial coordinates, and h denotes the state in temporal coordinates.

In these training samples, transfer information with replay experience is used as a collection of training samples, and it is ensured that each cohort is equipped with a complete sampling probability, the process is to ensure that the intelligences are all able to learn the associations between different states and behaviors more broadly, and to avoid overfitting to some specific experience sequences. The caching of data samples is done sequentially in the reward space with the cache D_t as shown in equation (4).

$$D_t = \{(e_1, e_2, \cdots, e_n)\} \tag{4}$$

Where: e_1 denotes the first data cache sample in the cache, and e_n denotes the nth data cache sample. After the intelligent caching of multiple threads, it is also able to complete the network gradient updating of each single thread on the basis of ensuring the cumulative gradient interaction, which is calculated as shown in Eq. (5).

$$a_x = a_i + \frac{Q_a - Y_s}{\alpha_p} \tag{5}$$

Where: α_x denotes the update metrics of the network gradient in a single thread, α_i denotes the cumulative selection probability of the actor, Q_a denotes the value of the initial function of the motion gradient, Y_s denotes the value of the termination function of the motion gradient, and α_p denotes the selection gradient metrics of state p.

The linear relationship between the initial function and the terminal function is shown in equation (6).

$$Q_a = \frac{h_k + Y_s}{h_k} \tag{6}$$

Where: h_k denotes the global network parameter after the gradient values have accumulated to a certain number of steps.

In the recognition of multi-intelligent body cache replay, the network update due to minimizing the loss function can be comprehensively expressed as a function of distinguishing work between the predicted and rewarded values, at this time, it is necessary to ensure the behavioral covariates of the global network by means of gradient descent as shown in Eq. (7).

$$F_k = \frac{\sum_{i=1}^{n} w_d y_i}{N_m} \tag{7}$$

Where: F_k denotes the behavioral parameters of gradient descent in the global network update; w_d denotes the weight attribute of the network update; y_i denotes the informative decision-making framework of the interaction equations; and N_m denotes the number of points of the skeletal points in the deep reinforcement learning, i.e., the number of the multi-threads after splitting them into the single threads.

Finally, the overall recognition result of the multithreading is calculated by the distance function, the

absolute value of which is shown in Equation (8).

$$D_{ij} = \frac{\left|\sqrt{(x_i - x_0)^2 - (y_i - y_0)^2}\right|}{x_i + y_i} \tag{8}$$

Where: D_{ij} denotes the relative distance of the multithreaded recognition state in the distance function, x_i and x_0 are the horizontal coordinates of the end point and the start point, and y_i and y_0 denote the vertical coordinates of the end point and the start point, respectively.

According to equation (8), it can be judged whether the recognition is successful or not, if the distance function is greater than or equal to 0, it means that the recognition is successful.

2.2. Multi-scale spatio-temporal feature encoder

In the human movement behavior recognition algorithm described in the previous section, the article constructs a human skeleton spatio-temporal movement model, and also introduces multi-intelligence deep reinforcement learning related methods to complete the efficient parallel recognition of students' behaviors in the sports classroom. In order to further improve the adaptation level of the model to complex sports scenes, the study designs a multi-scale spatio-temporal feature encoder, which employs a separated spatio-temporal pooling attention mechanism to capture the changes of spatio-temporal features at different scales.

The spatio-temporal features at different scales are extracted progressively based on the attention mechanism in the multiscale separated spatio-temporal feature encoder. Since computing the global self-attention in both temporal and spatial dimensions simultaneously brings computational cost in the video task, thus the separation of temporal and spatial attention is adopted to reduce the computational amount effectively while the performance is also comparable. Different from the traditional attention mechanism, this paper designs the Separated Spatio-Temporal Pooled Attention (SSTPA) module, which contains two sub-modules, Temporal Pooled Attention (TPA) and Spatial Pooled Attention (SPA), which introduces the pooling operation prior to the computation of the attention, and is able to reduce the computational amount further while extracting the multi-scale spatio-temporal features. The design details in each module in the encoder are described in detail below.

2.2.1. Separation of spatio-temporal pooling of attention

Separate spatio-temporal pooling attention consists of two parts, SPA, the spatial pooling attention module, and TPA, the temporal pooling attention module. Let the input, $\hat{X} \in \mathbb{R}^{L \times D}$, be a sequence of vectors containing L vectors of dimension D, where $L = T \times H \times W$.

For the spatial pooling attention module as shown in Fig. 2, is to compute the attention between all the vector sequences in the same frame.

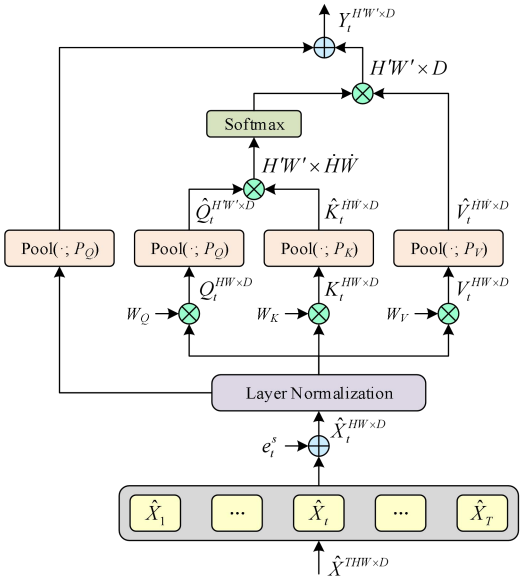


Figure 2. Spatial Pooling Attention Module.

where $\oplus$ denotes the tensor addition operation, $\otimes$ denotes the tensor multiplication operation, and the upper right corner label of the variable denotes the dimension of the feature. First take all the input vectors $\hat{X}_t \in \mathbb{R}^{HW \times D}$ in the t th frame, and map them into $Q_t / K_t / V_t$ by linear transformation, which can be expressed as:

$$Q_t = W_{Q_t} LN(\hat{X}_t + e_s) \tag{9}$$

$$K_t = W_{K_t} LN(\hat{X}_t + e_s) \tag{10}$$

$$V_t = W_{V_t} LN(\hat{X}_t + e_s) \tag{11}$$

where LN denotes layer normalization, $W_{Q_t} / W_{K_t} / W_{V_t}$ are all learnable tensors of dimension $D \times D$, and e_s are spatial location variables.

Different from the classical self-attention mechanism, in order to extract multi-scale spatial features while reducing the computation, the obtained $Q_t / K_t / V_t$ is first subjected to a pooling operation, i.e., downsampling, and the pooling operator is denoted as $Pool(\cdot, P)$, with P being the pooling kernel scale of the pooling operation, and commonly used pooling methods are maximum pooling, maximum pooling, and maximum pooling. P is the pooling kernel scale of the pooling operation, and the commonly used pooling methods are maximum pooling, average pooling and convolution. In order to ensure that the dimensionality is consistent, the pooling scale for K and V should be the same, and Q can choose the pooling scale according to the actual situation, that is to say, it satisfies the pooling kernel scale into $P_K = P_V$. The pooling operation for $Q_t / K_t / V_t$ can be expressed as:

$$\hat{Q}_t = Pool(Q_t; P_Q) \tag{12}$$

$$\hat{K}_t = Pool(K_t; P_K) \tag{13}$$

$$\hat{V}_t = Pool(V_t; P_V) \tag{14}$$

Then the attention of $\hat{Q}_t / \hat{K}_t / \hat{V}_t$ is calculated after pooling as shown in equation (15). where $\frac{1}{\sqrt{d}}$ scale transformation factor and SM denotes the softmax activation function.

$$Attention(Q_t, K_t, V_t) = SM\left(\frac{\hat{Q}_t \hat{K}_t^T}{\sqrt{d}}\right)\hat{V}_t \tag{15}$$

Finally, the output after computing the attention is kept residually connected to the input $\hat{X}_t$. Due to the pooling operation before computing the attention, to keep the dimensionality consistent, the residual connection here is a pooled residual connection, i.e., the inputs are pooled by doing the same pooling operation as Q and then summed up, i.e.,:

$$Y_t = Attention(Q_t, K_t, V_t) + Pool\left(LN\left(\hat{X}_t + e_s\right); P_Q\right) \tag{16}$$

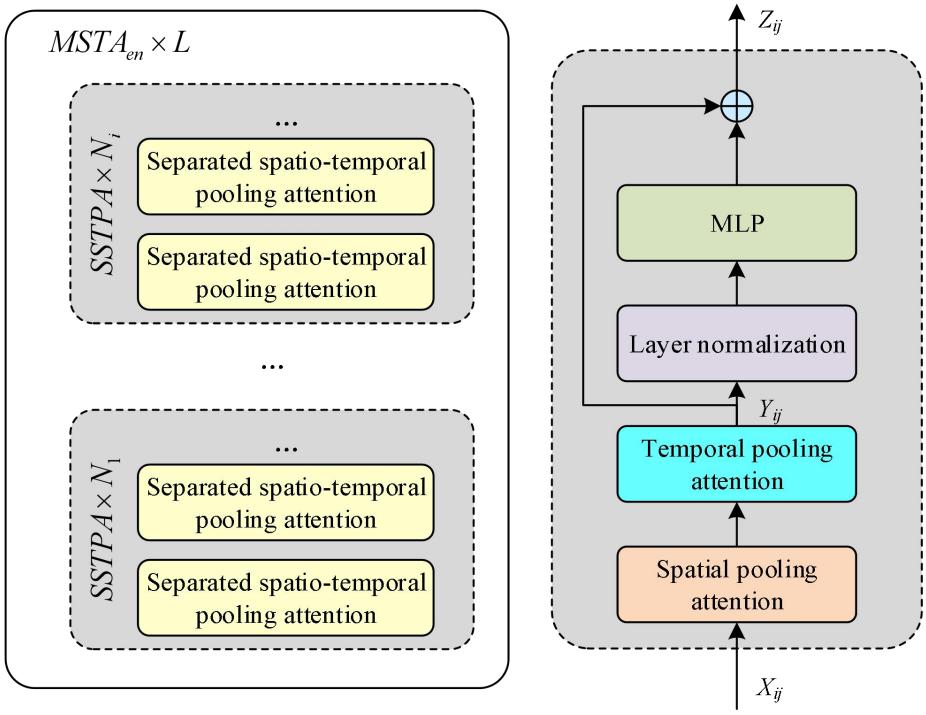
The pooled attention between all frames is computed sequentially to obtain the output $Y \in \mathbb{R}^{L' \times D}$ of the spatial pooled attention module, which is also the input to the temporal pooled attention module, where $L' = T \times H' \times W'$, H' & W' are spatially pooled after scaling. Temporal pooling attention is similar to spatial pooling attention in that it computes the attention between a sequence of vectors Y_s between different frames at the same spatial location, and the pooling operation acts on the temporal dimension, which can be summarized in Eq. (17), with the output being $Z \in \mathbb{R}^{L'' \times D}$, where $L'' = T' \times H' \times W'$, T' the pooled time dimension. The computational details of the temporal pooling attention module are consistent with the spatial pooling attention and will not be repeated.

$$Z_s = SM\left(\frac{Pool(Q_s;P_Q)Pool(K_s;P_K)^T}{\sqrt{d}}\right)Pool(V_s;P_V) + Pool(LN(Y_s+e_t);P_Q) \tag{17}$$

where $Q_s = W_{Q_s}LN(Y_s+e_t)$, $K_s = W_{K_s}LN(Y_s+e_t)$, $V_s = W_{V_s}LN(Y_s+e_t)$, Y_s is all vectors in the dimension when they are located at position s, and e_t is the temporal position vector. Note that P_Q, P_K, P_V are not the same as in the spatial pooling attention module, but instead pooling operations are performed in the time dimension.

2.2.2. Multi-scale spatio-temporal feature encoder

In a multiscale spatio-temporal encoder, its internal spatio-temporal features $F \in R^{T\times H\times W\times D}$ are showing an asymptotic change: the model initially focuses on localized detailed features, and thus possesses finer temporal and spatial features, and at the same time the number of feature channels is smaller, i.e., $T/H/W$ is larger and D is smaller; The later spatio-temporal features are more abstract and have a large number of feature channels, and are more concerned with global features, i.e., smaller $T/H/W$ and larger D. Such spatio-temporal scale variation is also more similar to the feature extraction process in convolutional neural networks. The structure of the multiscale spatio-temporal encoder is shown in Fig. 3(a), which consists of L multiscale spatio-temporal attention (MSTA) modules stacked together, with different MSTAs having different spatio-temporal feature scales and the same spatio-temporal feature scales in the same MSTA. MSTA contains N_i separated spatio-temporal pooling attention SSTPA modules, where $i = 1,2,\ldots,L$. Generally different MSTA modules contain different numbers of SSTPA modules, and the SSTPA structure is shown in Fig. 3(b).



(a) Encoder architecture

(b) Separated spatio-temporal pooling attention module structure

Figure 3. Multiscale Spatiotemporal Feature Encoder.

For the input X_{ij} of the j th SSTPA module in the i th MSTA, its SSTPA module output Z_{ij} is

computed as follows:

$$Y_{ij} = TPA\left(SPA(X_{ij})\right) \tag{18}$$

$$Z_{ij} = MLP\left(LN(Y_{ij})\right) + Y_{ij} \tag{19}$$

where SPA denotes spatially pooled attention, TPA denotes temporally pooled attention, and MLP is a multilayer perceptron, i.e., a fully connected layer.

In MSTA, in order to ensure the consistency of the spatio-temporal feature scale and fully learn the spatio-temporal features under this scale, thus only the downsampling of temporal and spatial dimensions is carried out in the first SSTPA module therein, and subsequent SSTPAs carry out the pooling operation with the pooling kernel of 1×1, thus maintaining the consistency of the feature scale with that in MSTA.

2.3. Graph Convolutional Adaptive Enhanced Network Architecture

In order to integrate the aforementioned features and achieve dynamic optimization, the article proposes the graph convolution adaptive augmentation network architecture DT-AGCN, which achieves adaptive augmentation recognition of students' behavioral patterns through temporal dimensionality transformation (T-DT) and graph structure modeling.

2.3.1. Network Architecture Diagram

The adaptive graph convolutional layer is able to co-optimize with other parameters in the network through an end-to-end learning mechanism, thus ensuring that the topology of the graph exhibits uniqueness across layers and samples. This flexibility allows the model to exhibit higher adaptability and robustness when dealing with complex data.

In previous GCN-based approaches, graph convolution operations are usually performed on skeleton data relying on ST-GCN predefined human body graphs. However, instead of directly utilizing convolutional neural networks (CNNs) to process graph data in real-world applications, these methods rely on predefined topologies, which limits their performance and flexibility when facing more complex and variable tasks. Therefore, the DT-GCN model constructed in this paper proposes to advance the development of action recognition models by improving the graph convolution approach to better capture the dynamic properties of skeleton data, especially the processing of time series information. In practice, graph convolution does not directly use convolutional neural networks to process graph data. To programmatically implement a GCN, the spatial graph convolution operation can be described using the following equation:

$$f_{out} = \sum_{k}^{K_v} W_k (f_{in} A_k) \tag{20}$$

where $f_{in} \in R^{C_{in}\times T\times N}$ is the input feature and $f_{out} \in R^{C_{out}\times T\times N}$ is the output feature where N and T denote the number of skeleton joints and frames included in the action samples, respectively, K_v denotes the spatial convolutional kernel size while C_{in} and C_{out} denote the input and output channels, respectively. $W_k \in R^{C_{out}\times C_{in}\times 1}$ is the weight matrix of the 1×1 convolution. $A_k = \Lambda_k^{-\frac{1}{2}} \overline{A}_k \Lambda_k^{-\frac{1}{2}}$. $\overline{A}_k \in R^{N\times N}$ Natural connections existing between joints can be obtained by computing the adjacency matrix and summing through the unit matrix of the self-connection matrices between joints. The Λ_k is a regularized diagonal matrix given by $\Lambda_k^{ij} = \sum_j (\overline{A}_k^{ij}) + \alpha$, where α is set to 0.001 avoids pure zeros in matrix operations.

Given that the fixed graph form of skeleton data is not suitable for skeleton-based action recognition, in this regard, the article continues to propose a transformed temporal sequence network (DT-AGCN) model, which constructs an adaptive graph convolutional layer using a fixed graph with variable graphs to sense to be better adapted to the task of action recognition, and the network architecture of the DT-AGCN model is shown in Fig. 4.

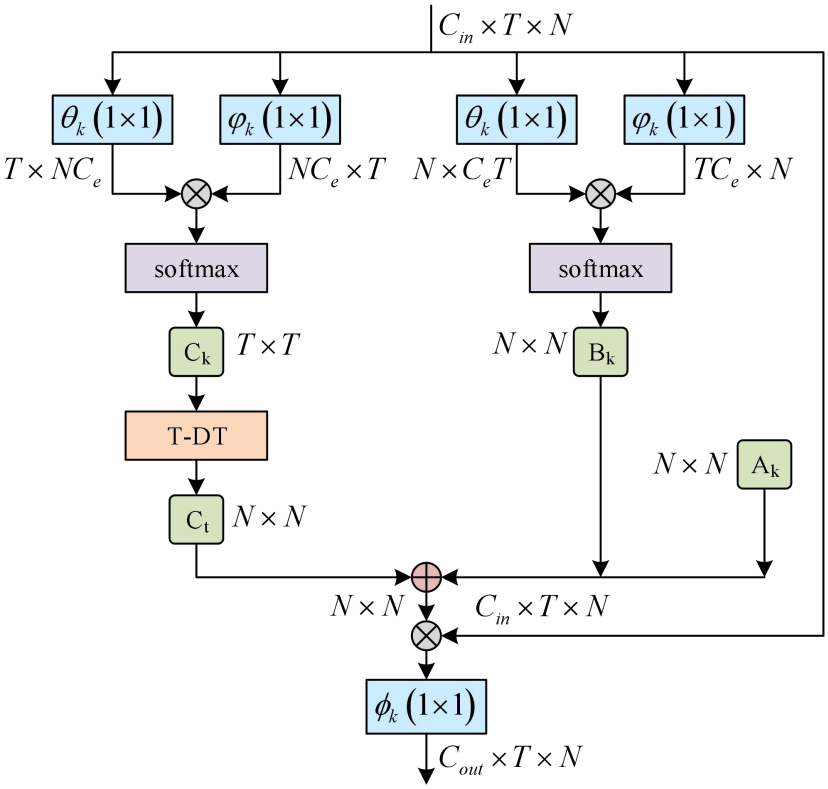


Figure 4. DT-AGCN.

The model can create adaptive graph structures based on different times and channels and create the final multidimensional adaptive graph for the DT-AGCN module by summing the adaptive graphs used for both with the initial adjacency matrix and the parameterized adjacency matrix, which allows the model to capture complementary discriminative features containing multidimensional data. The adaptive graph structure is generated by modifying Eq. (20) as shown below:

$$f_{out} = \sum_{k}^{K_v} W_k f_{in}(A_k + B_k + C_t) \tag{21}$$

The key difference between Eqs. (20) and (21) is that both are reflected in the adjacency matrix and it is divided into three parts: A_k, B_k and C_t.

The first subgraph, A_k, shows the global graph structure obtained through data learning and is intended to provide a more appropriate graph topology for the action recognition task. The adjacency matrix of this graph was initially constructed based on a human model, where elements were treated as parameterizable variables and updated together with other parameters during the training process. During this process, the value of A_k is not constrained in any way, implying that the structure and properties of the graph can be adaptively adjusted based solely on the training data. Through this data-driven approach, the model is able to effectively learn the feature graphs that are highly relevant to the recognition task. Notably, the specific value of A_k is unique at each layer, which allows the semantic information introduced at different layers to be more personalized to fit the complexity and diversity of the data being processed. This feature not only improves the flexibility of the model but also enhances its performance in the action recognition domain.

The second parameter B_k is learned according to the change of spatial dimensions and thus learns to get a graph for each sample, and in this paper, we use a normalized embedded Gaussian function to compute the similarity of two joints:

$$f(v_i, v_j) = \frac{e^{\theta(v_i)^T \varphi(v_j)}}{\sum_{j=1}^{N} e^{\theta(v_i)^T \varphi(v_j)}} \tag{22}$$

where N denotes the number of vertices.

In order to evaluate the similarity of each pair of vertices in the embedding space, the dot product operation is used in this case. In particular, the input feature mapping is processed by the embedding functions θ and φ before being fed into the embedding space, where $f_{in} \in R^{C_{in} \times T \times N}$ is the input feature, and the space after embedding is $R^{C_e \times T \times N}$. This paper is based on the previous research and experiment accordingly, so the embedding function in the method uses 1×1 convolution. The $N \times N$ spatial dimensional similarity matrix is then created by multiplying the two reconstructed embedding feature mappings, which are now denoted as the $W_\theta \in R^{C_e \times C_{in} \times 1}$ matrix and the $W_\varphi \in R^{C_e \times C_{in} \times 1}$ matrices. Since the matrix values are normalized to 0-1 and combined with the softmax activation function, this paper can be calculated according to the following equation.

$$B_k = soft\max\left(f_{in}^T W_{\theta k}^T W_{\varphi k} f_{in}\right) \tag{23}$$

where $W_{\theta k}$ and $W_{\phi k}$ denote the parameters of the corresponding embedding function, respectively.

The third parameter C_t has a unique graph for each sample, which helps to understand the particular interactions between joints based on the temporal information contained in the various motion samples. More precisely, in order to evaluate the feature similarity between the motion samples of each two frames in the temporal dimension, in this paper, we first use two normalized embedding Gaussian functions, which are similar to the adaptive methods used to evaluate the feature similarity between arbitrary pairs of joints in the spatial dimension. A set of time-dimensional embedded feature mappings obtained by processing the Gaussian functions are then transformed into $T \times C_e N$ and $C_e N \times T$, respectively. The time-dimensional similarity matrix is then generated by multiplying these two embedded feature mappings with the following formula:

$$C_k = soft\max\left(f_{in}^T W_{\theta k}^T W_{\varphi k} f_{in}\right) \tag{24}$$

Finally, they are combined through residual linking to enhance the multi-scale representation of the model.

2.3.2. T-DT Timing Dimension Transformation

In order to construct a temporal adaptive graph C_t, a T-DT temporal dimension transformation module is proposed, which is designed to efficiently process the temporal similarity matrix $T \times T$.The T-DT temporal dimension transformation process is shown in Fig. 5.

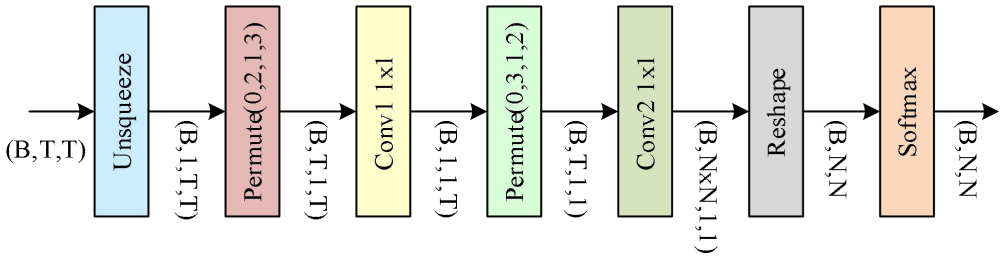


Figure 5. T-DT Time-Domain Transformation Process.

First, for a given temporal similarity matrix of the middle layer, the initial dimension of this matrix is (B,T,T) where B denotes the batch size and T denotes the number of time frames. In this paper, we first reconstruct this matrix by changing its shape from (B,T,T) to $(B,1,T,T)$, in which a new dimension is introduced so that the paper can process the temporal information in more detail.

Next, this paper utilizes a 1×1 convolutional layer named Conv1 to perform a squeezing operation on the first temporal dimension of each moment to extract its potential features. This process actually effectively reduces the redundant information of the time dimension through spatial compression of the convolution operation. Subsequently, in this paper, the other time dimension is considered as a convolutional channel, and the T -dimensional vector is transformed into a $N \times N$ matrix through a

single 1×1 convolutional layer named Conv2. This stage of transformation not only enhances the feature characterization, but also ensures the computational efficiency of the subsequent steps. Eventually, the temporal similarity matrix $T\times T$ is successfully transformed into a temporal adaptive graph C_t with dimension (B,T,T). This shape is designed so that C_t can be consistent with the dimensions of other feature graphs A_k, B_k and thus provides the necessary conditions for the subsequent information fusion. This processing method not only optimizes the flow of information, but also lays the foundation for deeper feature learning, which provides a strong guarantee for the performance improvement of the model.

3. Experimental validation and analysis of students' behavioral pattern recognition in physical education teaching scenarios

After constructing the student behavior pattern recognition model based on the adaptive algorithm, this chapter proceeds to the systematic experimental validation and analysis in order to comprehensively consider the effectiveness and robustness of this series of methods in the real sports teaching environment. Firstly, the model's ability to understand the semantics of basic sports actions is verified at the level of local behavior-object interactions, and then the model's ability to recognize complete sequences of continuous actions is evaluated in two video datasets, KTH and UCF, with a focus on its ability to adapt to changes in complex teaching environments.

3.1. Overview of the data set

In order to verify the application of the human movement behavior recognition model based on adaptive algorithm proposed in this paper in specific sports teaching, experiments are conducted on HICO dataset, KTH dataset and UCF dataset respectively.

The HICO dataset is a sports human-object interaction image dataset. The dataset has a total of 27,398 images, which cover the combination of 57 common interaction behaviors (e.g., holding, passing, shooting, padding, swinging, bracing, etc.) and 25 types of sports equipment (e.g., basketballs, volleyballs, ping-pong paddles, gymnastics mats, vaulting boxes, jumping ropes, etc.) in the sports classroom, constituting 406 sports classroom-specific human-object interaction categories, such as holding basketballs for breakthroughs, padding volleyballs for preparation, and bracing on vaulting box support, etc. This dataset retains the multi-label characteristics of real-world scenarios, which facilitates the model's understanding of local behavioral co-occurrences in complex, combinatorial sports actions.

The KTH dataset contains 629 videos focusing on 10 basic and assessed movements in physical education, namely: running, high leg running, lunge press, standing long jump, forward roll, radio gymnastics, shooting, seated forward bending, plank support, and jumping rope. These maneuvers were performed by multiple students of different ages in four different physical education environments, C1 (standard playground environment), C2 (indoor gymnasium environment), C3 (spectator viewpoint environment in the presence of partial occlusion), and C4 (outdoor environment in different weather).

The UCF dataset contains 178 motion videos filmed in real primary and secondary school physical education classes, covering nine types of typical physical education activity behaviors: fast running, walking, playing basketball, playing soccer, playing badminton, playing table tennis, playing volleyball, gymnastics, and jumping rope. The videos in this dataset were shot in a real playground and gymnasium with extremely noisy backgrounds, with low uniformity of students' attire, and included a lot of dynamic background and lighting changes caused by students' movement and camera following. This dataset is used to comprehensively evaluate the model's ability to recognize and differentiate collective, confrontational, and free-activity behaviors in real, open, and complex physical education scenarios.

3.2. Local Behavioral Research Analysis

In the HICO dataset, 80% of each of these categories were used as the training set and the rest as the test set. A total of 73 local human behavior labels were obtained in this section, including "hip: sit", "hand: hold", "foot: jump", etc.

The number of co-occurrences between different local behaviors/objects is calculated on this dataset, followed by constructing a co-occurrence statistics matrix and a co-occurrence probability matrix sequentially based on this statistical information. In the above process, the size of the constructed matrices is 98×98 because 73 local behaviors and 25 devices were obtained from the training data.To describe the process more clearly, 10 local behaviors and 5 sports devices were extracted from the two matrices for display due to the limitation of the page space.The statistical matrix of the number of times of co-occurrences of local behaviors/objects for the HICO dataset is shown in Fig. 6. (where B denotes

behaviors and O denotes objects, B1 is the first interactive behavior “hand: hold” and O1 is the first object “vault”).

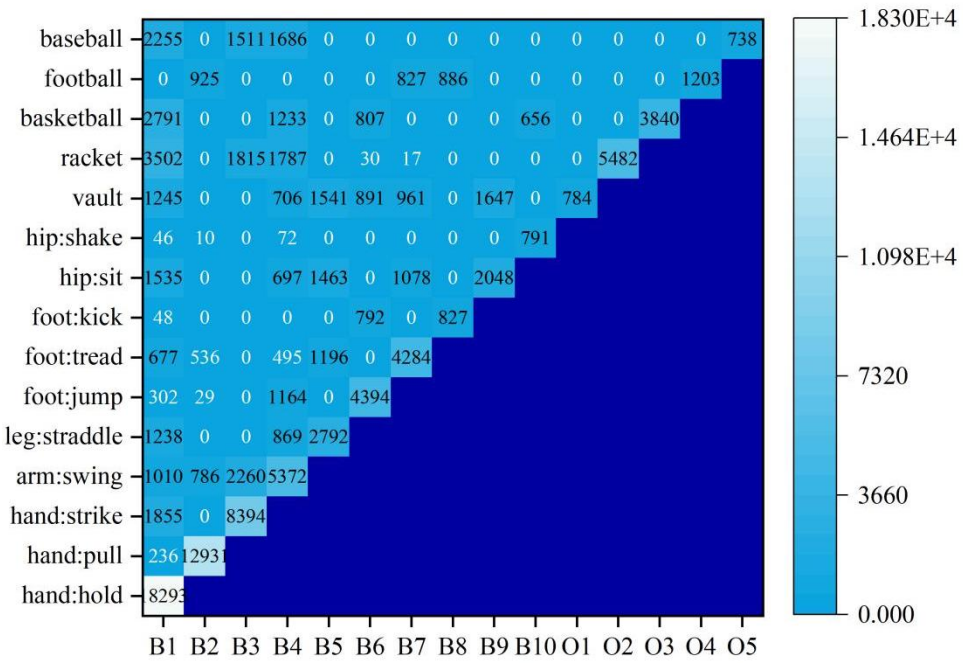


Figure 6. The local behavior/object co-occurrence frequency matrix.

It can be seen from the co-occurrence frequency statistical matrix that the numerical distribution shows a relatively obvious aggregation, and there are extensive and strong co-occurrence relationships between local behaviors and between local behaviors and objects. The basic behavior "hand: holding" and various equipment all show a high-frequency co-occurrence phenomenon. For instance, it appears 3,502 times with "racket", 2,791 times with "basketball", and 2,255 times with "baseball". This clearly indicates that "holding" is the fundamental interaction method for most equipment sports. There is also a clear synergy among the behaviors. For instance, "arm: swing" and "hand: strike" occur together 2,260 times, establishing a typical action framework for swinging the arm to hit the ball. The combination of "sitting on the hips" and "crossing legs" occurred 1,463 times, which is consistent with the postural characteristics of equipment preparation movements such as the vault.

The probability values provide a clearer picture of the conditional chances of another behavior or object appearing at the same time as a particular behavior or object, and the local behavior/object co-occurrence frequency matrix in the HICO dataset is shown in Figure 7.

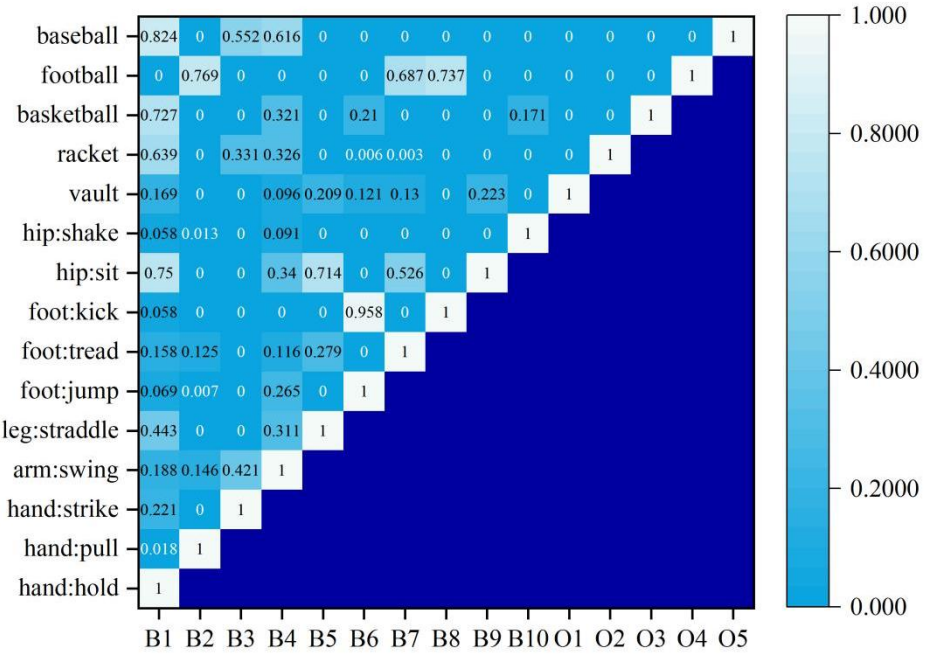


Figure 7. The local behavior/object co-occurrence probability matrix.

When a "baseball" appears, the probability of the "hand:" holding "action is as high as 82.36%, while the probability of the" hand: hitting "action is also 55.19%, indicating that when a baseball appears, it is usually accompanied by holding and hitting actions. Another typical case," football ", has a probability of appearing together with "foot: kicking" as high as 73.65%. Data analysis also identified some high-co-occurrence combinations that conform to the human body's coordination mechanism but seem rather unconventional. The simultaneous occurrence probability of "foot: pedaling" and "buttocks: sitting" was 52.64%, possibly due to the repeated appearance of the buttocks sitting while cycling in coordination with the pedals in the dataset videos.

3.3. Research on motor action recognition in physical education

The above experiments conducted for the HICO dataset fully validate the ability to understand the semantics of human-object interactions based on the adaptive algorithm model, and in order to further measure the ability of the adaptive algorithm proposed in this paper in recognizing complete sequential actions, this section implements more specific sports action recognition experiments on the two datasets of KTH and UCF.

3.3.1. Analysis of experimental results for the KTH dataset

Firstly, on the KTH dataset, the model's performance in complex scenarios of real teaching is comprehensively analyzed in terms of two dimensions: action classification accuracy and different scenarios.

The confusion matrix of various action behaviors of the model on the KTH dataset is shown in Figure 8.

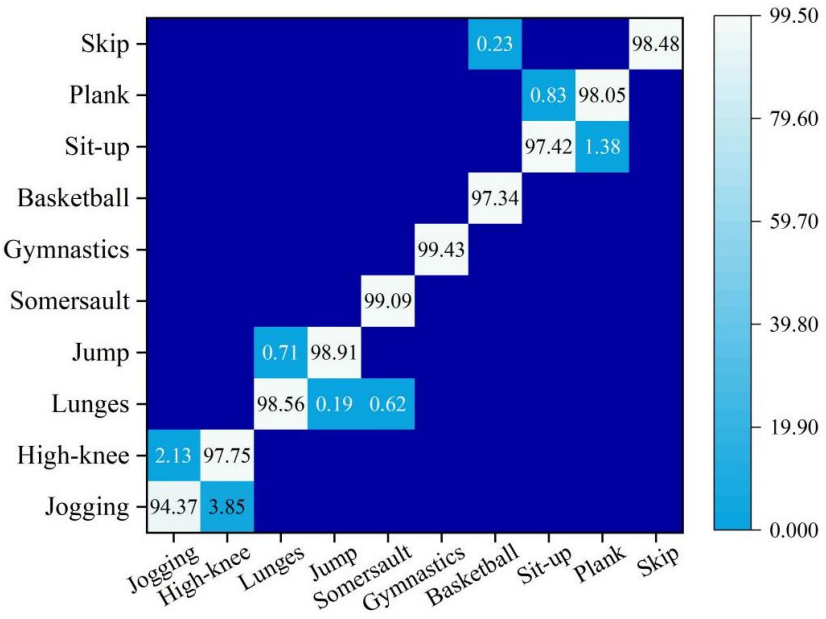


Figure 8. Confusion matrix of various action behaviors on the KTH dataset.

The adaptive behavior recognition model proposed in this paper demonstrates excellent recognition performance for ten types of basic physical education teaching actions. The recognition accuracy of each action exceeds 97%. The model has the highest accuracy in recognizing "standing long jump", "forward roll" and "rope skipping", reaching 98.91%, 99.09% and 98.48% respectively. This indicates that the model has an extremely strong ability to distinguish periodic movements and large-scale dynamic actions with distinct spatiotemporal characteristics. It is worth noting that the model shows understandable mild confusion in some similar movements. For instance, there is a 1.38% misjudgment between "sit and reach" and "plank". This is probably because both exhibit the low dynamic feature of forward trunk tilt in terms of skeletal posture. The diagonal values of the vast majority of actions are extremely high, which effectively proves the effectiveness of the model in basic action recognition tasks based on the human skeleton spatio-temporal motion model and multi-scale feature encoders in actual physical education teaching activities.

In order to test the model's adaptability to environmental changes, it continued to be tested under each of the four different sub-scenes (C1, C2, C3, and C4) of the KTH dataset, and the recognition accuracies of the different actions are shown in Figure 9.

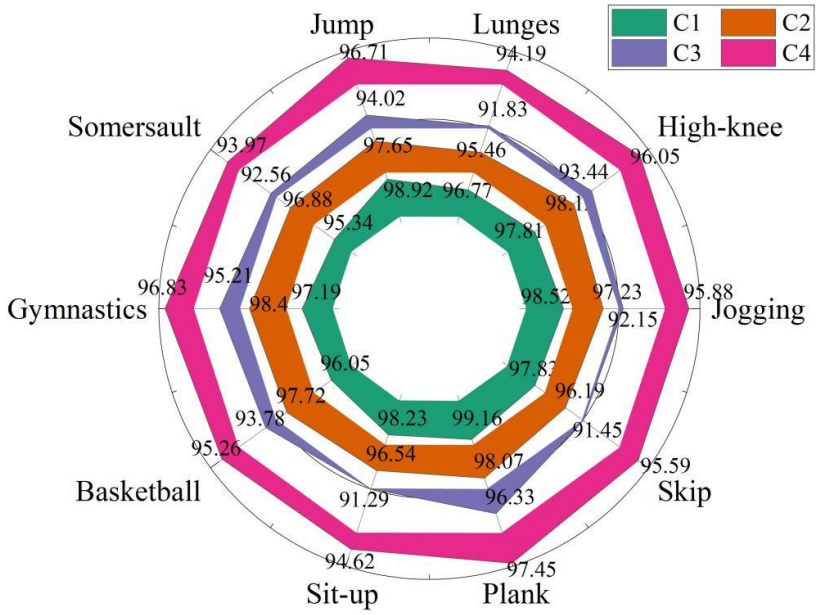


Figure 9. The recognition precision of each action in 4 different scenarios.

The model demonstrated outstanding performance in all test scenarios. The average accuracy rates for the C1 standard playground, C2 indoor gymnasium, and C4 outdoor variable weather conditions were 97.68%, 97.23%, and 95.75% respectively. Even in the most challenging C3 partial occlusion scenario, the recognition rate remained at 93.11%. In terms of stable postures and clear trajectories such as "plank" and "standing long jump", the model demonstrates top-notch recognition accuracy. However, for low-dynamic and easily masked movements such as "lunge leg presses" and "sit and reach", the recognition rate is relatively lower, but it still remains above 90%. In the indoor gymnasium environment, the recognition rates of radio calisthenics and high leg lifts reached 98.4% and 98.12% respectively, slightly higher than that in the standard playground environment. This is mainly because the uniform lighting and stable background create favorable conditions for the model to focus on the movement features. Data shows that the multi-scale spatio-temporal feature encoder and graph convolution adaptive augmentation network built by the research institute can effectively address the problems of lighting changes, weather disturbances and local occlusion, meeting the application requirements of complex physical education teaching scenarios.

3.3.2. Analysis of experimental results on the UCF dataset

In order to further validate the generalization ability and practicality of the model when it is more relevant to real teaching environments, the experiments in this section will be conducted on the more challenging UCF dataset. This dataset contains long videos of real classrooms with noisy backgrounds, strenuous camera movements, and significant differences in students' attire and behavioral styles, and the model's confusion matrix for recognizing nine types of typical physical education behaviors in the UCF dataset is shown in Fig. 10.

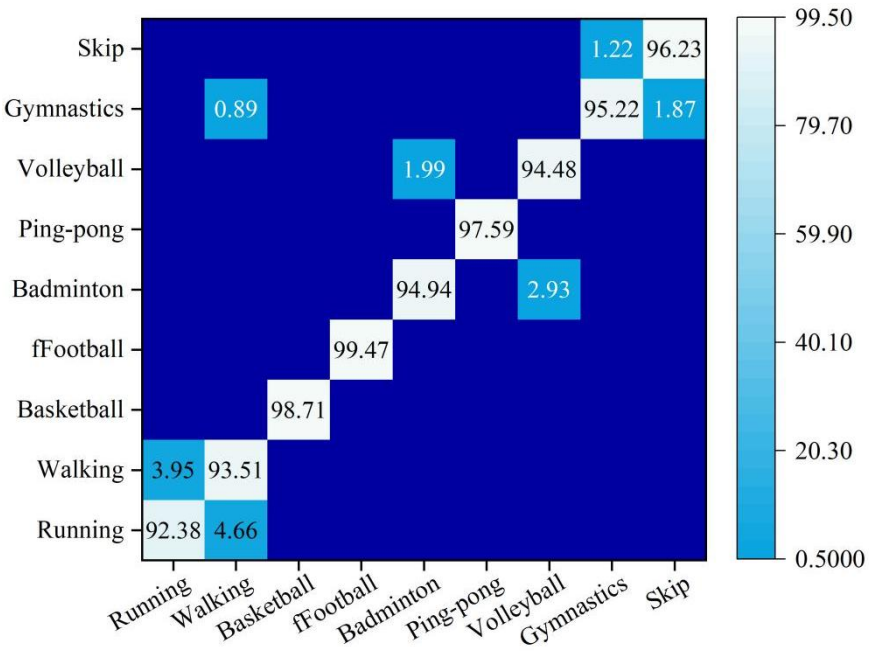


Figure 10. Confusion matrix of 9 typical sports teaching activity behaviors.

The model has an extremely strong recognition ability for confrontational sports such as "playing football" and "playing basketball", which present distinct spatiotemporal patterns and have strong interactivity. The recognition rates are 99.47% and 98.71% respectively, and there is no confusion between them. It also demonstrated excellent performance in small-scale, rapid and repetitive actions such as "playing table tennis" (97.59%) and "skipping rope" (96.23%). In some categories with similar visual features and severe background interference, there have been plausible confusions, mainly manifested as approximately 4% mutual misjudgment between "running" and "walking", which is due to the similar characteristics of their movement trajectories. There is about 2-3% confusion between "playing badminton" and "playing volleyball" because the movements of swinging arms to hit the ball are very similar. Both "gymnastics" and "rope skipping" contain elements of jumping, resulting in approximately 1.87% incorrect judgments. The recognition rate of all action dominant diagonals exceeds 92%. Meanwhile, the model as a whole still maintains an average accuracy rate of as high as 95.84% in real, open and dynamic teaching scenarios, fully demonstrating its high practicality in more complex

actual physical education teaching applications.

4. Application of Behavioral Pattern Recognition Models Based on Adaptive Algorithms

In order to comprehensively test the effectiveness and application effect of the student behavior pattern recognition model based on adaptive algorithm constructed in this paper in actual physical education teaching, the model is integrated into a real teaching environment, and a semester-long teaching empirical study is implemented, taking 263 students from the 2024 session of the College of Economics of a university as the survey object, and adopting the arrangement of the intelligent teaching system with real-time movement posture analysis and recognition capability, the process data collection and monitoring of student behavior performance in the physical education classroom are continuously and objectively conducted. Continuously and objectively collect and monitor students' behavioral performance in the physical education classroom.

Focusing on the four core dimensions of sports participation, skill attainment, physical fitness pass rate, and classroom interaction frequency, the 12-week tracking study was conducted to evaluate the effects of the model-supported physical education teaching mode on students' learning process and effects.

4.1. Overall change in dimensional monitoring

The changes in students' dimensions from week to week in this physical education model are shown in Table 1.

Table 1. The changes of students in each week under the sports teaching model.

	Exercise participation rate/%	**Skill attainment rate/%**	**Physical fitness pass rate/%**	**Frequency of classroom interaction**
Week1	58.51	52.43	62.83	26.8
Week2	61.00	63.49	69.91	24.3
Week3	60.45	63.22	75.55	33.6
Week4	68.89	65.16	84.52	33.3
Week5	73.46	69.44	88.51	39
Week6	76.37	75.40	92.27	50.1
Week7	81.21	80.39	95.37	50.3
Week8	86.20	85.51	96.27	56.8
Week9	89.52	86.20	97.36	60.9
Week10	92.28	91.59	98.37	71
Week11	95.89	94.08	98.47	70.9
Week12	98.93	97.54	99.03	72.3

In terms of the overall trend, all of the observed indicators show a continuous and steady increase, indicating that this teaching mode has positive and comprehensive positive effects on students. At the beginning of the semester, the students' indicators were at a moderate or low level, with only 58.51% of students participating in sports, 52.43% of students meeting the skill requirements, 62.83% of students passing physical fitness, and an average of 26.8 interactions in the classroom, which is a common situation under the traditional teaching model. As the instructional cycle progressed, the data climbed steadily, and by week 12, all dimensions were at a very high level, with sports participation and physical fitness passing rates nearly saturated at 98.93% and 99.03%, respectively, and skill attainment at 97.54%, with the frequency of classroom interactions rising to 72.3, a roughly 1.7-fold increase from the first week. The frequency of classroom interaction rose to 72.3 times, an increase of about 1.7 times compared with the first week. This series of data fully illustrates that the physical education teaching model based on the behavior recognition model of adaptive algorithm can effectively stimulate students' enthusiasm

for participation, greatly enhance the standardization and accuracy of the mastery of sports skills, improve the students' physical fitness level, and greatly promote the interactive and collaborative atmosphere of the classroom.

4.2. Trends in implementation process data

4.2.1. Trends in implementation process data

Trends in implementation process data in terms of campaign participation are shown in Figure 11.

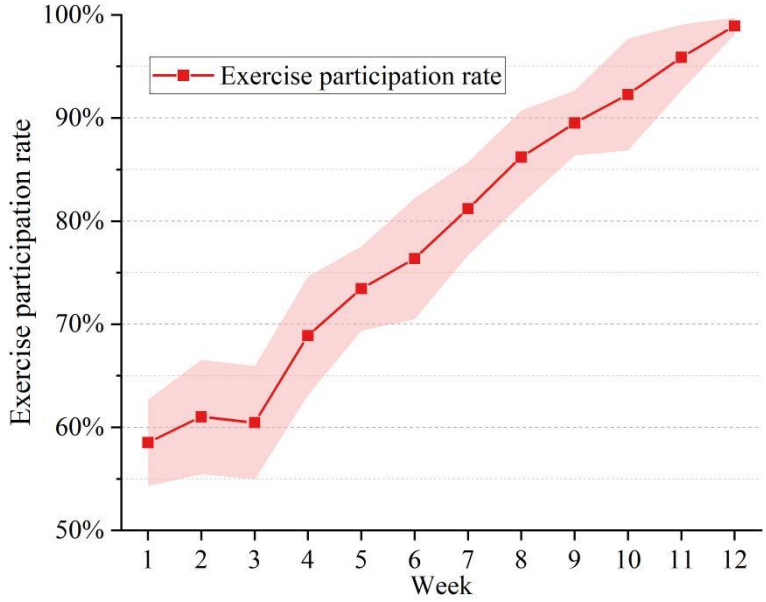


Figure 11. The trend of changes in the data on participation in sports.

Exercise participation of the 263 students who participated in the experiment showed a strong, nearly linear growth momentum throughout the experimental cycle. The growth was relatively smooth at the beginning, with clear signs of accelerated growth from week 3 onwards, and maintained at a high growth slope from week 8 onwards. This trend suggests that the immediate feedback and personalized goal setting given by the adaptive algorithm can effectively stimulate students' intrinsic motivation, allowing them to transition from passive participation to active engagement and reach a positive cycle of continuous reinforcement.

4.2.2. Skills attainment rate

Trends in implementation process data in terms of student physical skill attainment rates are shown in Figure 12.

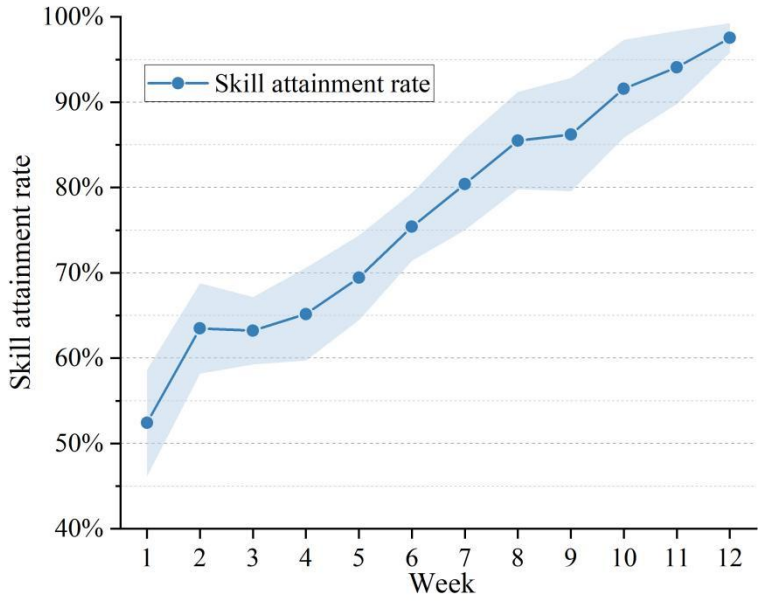


Figure 12. The trend of the data on the rate of skill attainment.

The change trend of the skill attainment rate shows a typical learning curve characteristic, the first two weeks of growth is faster, achieving a 63.49% attainment rate, indicating the obvious initial effectiveness of the new model in the immediate correction of movement specification. After the students adapted to the refined movement control, the attainment rate accelerated upward in the middle and late stages, and then always grew steadily in the later stages, which showed that the model played a precise role in assisting and optimizing the students' skill learning.

4.2.3. Physical fitness pass rate

Trends in implementation process data in terms of student physical education pass rates are shown in Figure 13.

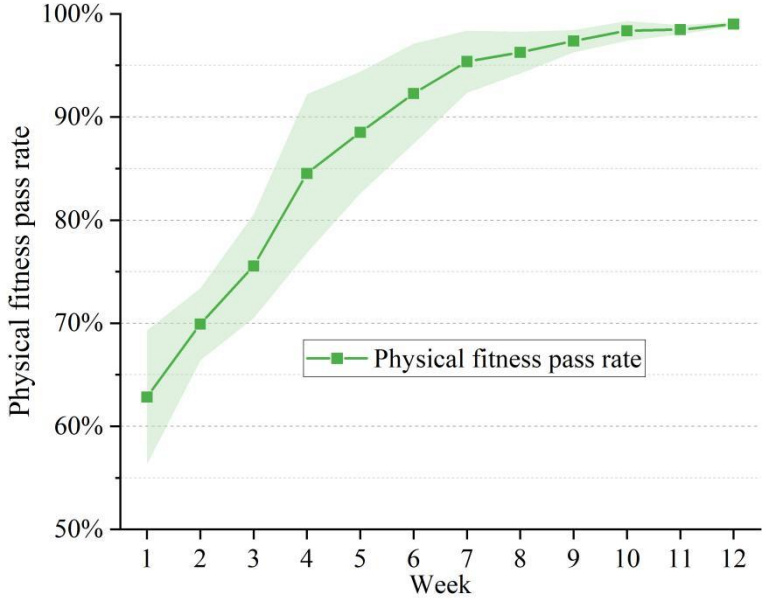


Figure 13. The trend of the data on physical fitness pass rates.

Over the 12-week experimental observation period, the physical fitness pass rate was the most rapid and steady improvement, showing strong growth from week 1 and remaining at a high level of over 90% from week 6 onwards. The growth curves indicate that the scientific and individualized exercise load training program developed by the model can efficiently accelerate the overall growth of students' basic

physical fitness, with obvious and long-term results.

4.2.4. Frequency of classroom interaction

Trends in implementation process data in terms of frequency of student classroom interactions are shown in Figure 14.

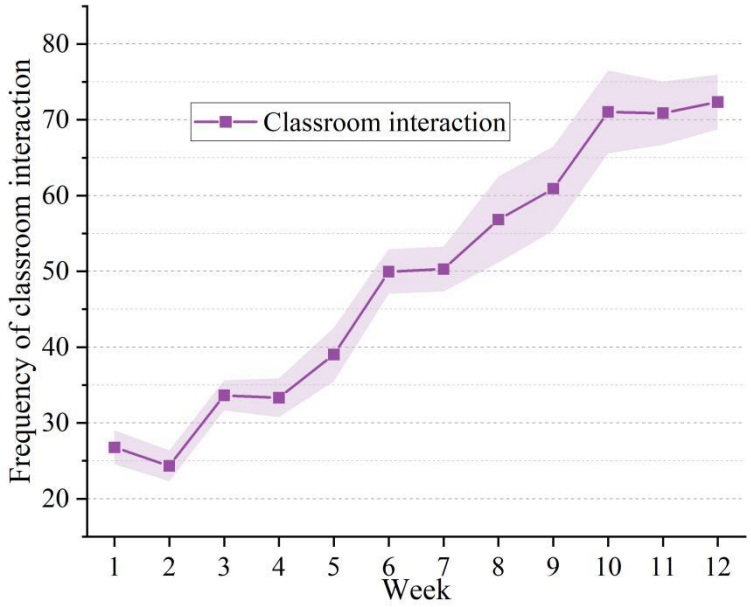


Figure 14. The trend of changes in the frequency of classroom interactions.

The direction of change in the frequency of classroom interactions reflects significant phase fluctuations and an overall rise, with a slight drop in the first period, perhaps related to the students' adaptation to the new teaching mode, and a rebound from the low level from the 3rd week onwards, showing signs of two significant leaps, and finally realizing the stabilization of classroom interactions at more than 70 per class period. This shows that the technical tools based on adaptive algorithms provide a backbone for designing more interactive teaching activities, successfully creating a lively and collaborative atmosphere in the classroom.

5. Conclusion

This paper focuses on the key technology of student behavior pattern recognition in physical education and its application, and constructs a set of recognition models based on adaptive algorithms.

In the HICO dataset that focuses on human-object interaction, the model successfully mined a strong co-occurrence relationship between local behaviors and sports equipment. The co-occurrence probability of "holding" and "racket" is as high as 63.88%, and the co-occurrence probability of "kicking with feet" and "football" is 73.65%.

(2) The recognition accuracy of the ten basic actions on the KTH dataset is as high as 98.12%, and more than 93% can still be maintained in complex scenes with partial occlusion.

(3) On the UCF dataset, which is closer to the real teaching situation, the overall recognition accuracy of the model reaches 95.84%, especially for the antagonistic sports of “playing soccer” (99.47%) and “playing basketball” (98.71%). The overall recognition accuracy is 95.84%.

(4) The intelligent teaching system equipped with the model greatly improves the quality of teaching and students' development, with the percentage of students' participation in sports increasing from 58.51% to 98.93%, the rate of skill attainment climbing from 52.43% to 97.54%, the pass rate of physical fitness increasing from 62.83% to 99.03%, and the frequency of classroom interactions increasing from the initial 27 times to 72 times.

The experiment fully verified that the behavior pattern recognition model based on adaptive algorithms can accurately identify behaviors, and can also effectively mobilize students to participate actively, correct technical movements, improve physical fitness and promote collaborative learning by virtue of real-time feedback and personalized interventions.

References

1. Eikeland, I., & Ohna, S. E. (2022). Differentiation in education: a configurative review. Nordic Journal of Studies in Educational Policy, 8(3), 157-170.
2. Haelermans, C. (2022). The Effects of Group differentiation by students' learning strategies. Instructional Science, 50(2), 223-250.
3. Geelan, D., Christie, P., Mills, M., Keddie, A., Renshaw, P., & Monk, S. (2015). Lessons from Alison: a narrative study of differentiation in classroom teaching. International Journal of Pedagogies and Learning, 10(1), 13-23.
4. Wang, Y., Li, T., Geng, C., & Wang, Y. (2019). Recognizing patterns of student's modeling behaviour patterns via process mining. Smart Learning Environments, 6(1), 26.
5. de Ruiter, J. A., Poorthuis, A. M., & Koomen, H. M. (2019). Relevant classroom events for teachers: A study of student characteristics, student behaviors, and associated teacher emotions. Teaching and Teacher Education, 86, 102899.
6. Tran, N., Nguyen, H., Luong, H., Nguyen, M., Luong, K., & Tran, H. (2023). Recognition of Student Behavior through Actions in the Classroom. IAENG International Journal of Computer Science, 50(3).
7. Li, J., & Xue, E. (2023). Dynamic interaction between student learning behaviour and learning environment: Meta-analysis of student engagement and its influencing factors. Behavioral Sciences, 13(1), 59.
8. Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (2015). Teaching behaviors, academic learning time, and student achievement: An overview. The Journal of Classroom Interaction, 50(1), 6-24.
9. Zhou, Y., Wang, J., & Zhang, J. (2024). A Multimodal Image Recognition System for Student Behavior Analysis in Smart Classrooms in Universities. Traitement du Signal, 41(6).
10. Rincon-Flores, E. G., Castano, L., Guerrero Solis, S. L., Olmos Lopez, O., Rodríguez Hernández, C. F., Castillo Lara, L. A., & Aldape Valdés, L. P. (2024). Improving the learning-teaching process through adaptive learning strategy. Smart Learning Environments, 11(1), 27.
11. Rincón-Flores, E. G., Mena, J., López-Camacho, E., & Olmos, O. (2019, October). Adaptive learning based on AI with predictive algorithms. In Proceedings of the seventh international conference on technological ecosystems for enhancing multiculturality (pp. 607-612).
    Zhu, J. (2021). Adaptive optimisation algorithm for online teaching behaviour. International Journal of Continuing Engineering Education and Life Long Learning, 31(3), 405-417.
12. Liu, C. (2025). Mining personalized learning feature labels from student reflection reports: A computational approach to adaptive education. International Journal of Mechanical Engineering Education, 03064190251375050.
13. Rahman, C. M., & Rashid, T. A. (2021). A new evolutionary algorithm: Learner performance based behavior algorithm. Egyptian Informatics Journal, 22(2), 213-223.
14. Tran, D. P., Nguyen, G. N., & Hoang, V. D. (2020). Hyperparameter optimization for improving recognition efficiency of an adaptive learning system. IEEE Access, 8, 160569-160580.
15. Li, F., He, Y., & Xue, Q. (2021). Progress, challenges and countermeasures of adaptive learning. Educational Technology & Society, 24(3), 238-255.
16. Barbosa, P. L. S., Carmo, R. A. F. D., Gomes, J. P., & Viana, W. (2024). Adaptive learning in computer science education: A scoping review. Education and Information Technologies, 29(8), 9139-9188.
17. Ezzaim, A., Dahbi, A., Aqqal, A., & Haidine, A. (2024). AI-based learning style detection in adaptive learning systems: a systematic literature review. Journal of Computers in Education, 1-39.
18. Roy, K., & Farid, D. M. (2024). An adaptive feature selection algorithm for student performance prediction. IEEE Access, 12, 75577-75598.
19. Popescu, D. A., Bold, N., & Stefanidakis, M. (2025). A Systematic Model of an Adaptive Teaching, Learning and Assessment Environment Designed Using Genetic Algorithms. Applied Sciences, 15(7), 4039.
20. Wu, Y. (2025, January). Adaptive Learning Algorithm in Artificial Intelligence Archives Data. In 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) (pp. 1-6). IEEE.
21. Jasim, A. H., & Hoomod, H. K. (2025, April). Intelligent student behavior recognition system for the classroom environment using hybrid deep learning. In AIP Conference Proceedings (Vol. 3282, No. 1, p. 030011). AIP Publishing LLC.
22. Huang, W., Li, N., Qiu, Z., Jiang, N., Wu, B., & Liu, B. (2020). An Automatic Recognition Method for Students' Classroom Behaviors Based on Image Processing. Traitement du Signal, 37(3).
23. Wu, B., Wang, C., Huang, W., Huang, D., & Peng, H. (2021). Recognition of Student Classroom Behaviors Based on Moving Target Detection. Traitement du Signal, 38(1).
24. Liu, L. (2024, December). Research on Student Behavior Recognition and its Application based on Machine Learning. In Proceeding of the 2024 International Conference on Diversified Education and Social Development (DESD 2024) (Vol. 899, p. 70). Springer Nature.
25. Zhang, H., & Li, Y. (2024). Student classroom teaching behavior recognition based on DSCNN model in intelligent campus education. Informatica, 48(9).
26. Yajun, W. (2025). A Deep Learning Recognition Method for Students' Abnormal Behaviors in Smart Classroom Teaching Scenarios. International Journal of High Speed Electronics and Systems, 34(04), 2540291.
27. Chonggao, P. (2021). Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm. Journal of Intelligent & Fuzzy Systems, 40(2), 2421-2431.

28. Yuvaraj, R., Prince, A. A., & Murugappan, M. (2024). An automated recognition of teacher and student activities in the classroom environment: A deep learning framework. IEEE Access.
29. Jisi, A., & Yin, S. (2021). A new feature fusion network for student behavior recognition in education. Journal of Applied Science and Engineering, 24(2), 133-140.
30. Qin, M. E. (2023, November). Research on Student Classroom Behavior Recognition Based on Computer Vision. In 2023 International Conference on Computer Simulation and Modeling, Information Security (CSMIS) (pp. 157-162). IEEE.