

International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556

<https://www.inderscience.com/ijris>

Application of convolutional neural network in normative detection and error correction of college students' physical exercise actions

Huan Chen

DOI: [10.1504/IJIRIS.2026.10076977](https://doi.org/10.1504/IJIRIS.2026.10076977)

Article History:

Received:	19 December 2025
Last revised:	20 January 2026
Accepted:	21 January 2026
Published online:	26 March 2026

Application of convolutional neural network in normative detection and error correction of college students' physical exercise actions

Huan Chen

School of Physical Education,
Ankang University,
Ankang, 725000, China
Email: c_1597@163.com

Abstract: To address the issues of experience-dependent movement standard evaluation and delayed error correction feedback in college PE, a bond product neural network-based method is developed for detecting and correcting students' exercise movements. Focusing on squats, push-ups, and lunges, the study builds a skeleton sequence dataset, and adopts a model combining convolutional feature extraction and multi-distribution attention weights for joint movement classification, standard scoring, and deviation correction. Experimental results show that the model's action classification accuracy on the test set is 92.7%–96.8%, with an average absolute error of 2.95 and a reasoning delay of ~11 ms. Normative scores are 77.6–84.1; 28.1% and 25.4% of corrections target insufficient squat depth and hip sinking. The results demonstrate that this method meets the accuracy and real-time demands of college PE classrooms, providing quantitative support for teaching evaluation and training intervention.

Keywords: coupon product neural network; college students; physical education in colleges and universities; normative score of action; error correction prompt.

Reference to this paper should be made as follows: Chen, H. (2026) 'Application of convolutional neural network in normative detection and error correction of college students' physical exercise actions', *Int. J. Reasoning-based Intelligent Systems*, Vol. 18, No. 10, pp.72–87.

Biographical notes: Huan Chen is a member of the School of Physical Education, Ankang University. He studied at Xi'an University of Architecture and Technology (2004–2008), Xi'an Physical Education University (2010–2013), and Valaya Alongkorn Rajabhat University (2020–2023).

1 Introduction

In college physical education, instructional effectiveness is often constrained by large class sizes, limited contact time, and the reliance on teachers' visual observation to evaluate movement quality. Under such conditions, instructors tend to focus on task completion rather than detailed movement execution, making it difficult to identify subtle but safety-relevant deviations such as insufficient squat depth or unstable joint alignment. These limitations reduce the precision of feedback and delay corrective intervention, which may weaken training outcomes and increase injury risk.

As physical education curricula increasingly emphasise standardised movement quality and individualised guidance, there is a growing demand for automated systems capable of continuously monitoring student actions and providing immediate, objective feedback. Rather than replacing instructors, such systems aim to support teaching by translating complex movement patterns into interpretable quantitative indicators that align with classroom instruction needs.

In a typical college physical education class, a group of students may repeatedly perform squats under limited supervision. While most students complete the required repetitions, subtle errors such as insufficient squat depth or unstable knee alignment often go unnoticed. Over time, these deviations reduce training effectiveness and may increase the risk of knee or lower back strain. Instructors, constrained by class size and time, are rarely able to provide immediate and individualised correction for every student. This common teaching scenario highlights the need for instructional support systems that can continuously monitor movement execution and provide timely, objective feedback. Addressing such practical teaching challenges serves as the primary motivation of this study.

In the field of intelligent algorithm research, the structural fusion of deep learning model has become the direction to improve the ability of time series prediction and complex pattern recognition. Shang et al. introduced attention mechanism into the combined structure of convolution and cyclic neural network around the task of financial time series prediction, and strengthened the modelling ability of key feature sections. Its research shows

that the hybrid network has a more stable performance when dealing with high noise and strong fluctuation data (Shang et al., 2024). The train of thought provides a reference for the modelling method of 'key segment prominent' in complex behaviour sequence.

In the research on sports and psychology in colleges and universities, Zhang et al. systematically combed the relationship among cognition, emotion and behaviour in college sports teaching from the perspective of sports psychology, and put forward that teaching intervention should pay more attention to psychological feedback and behaviour adjustment in the process of students' action execution (Zhang et al., 2024). Quan and Sun pointed out through empirical research that college students' behaviour adjustment ability is obviously different in the face of external events, and this difference is closely related to individual state variables (Quan and Sun, 2024). McCarron et al. started with the relationship between college students' labour participation and subjective experience, and emphasised that the observability and continuity of the behaviour process itself are of great value for evaluating individual status (McCarron et al., 2025). These studies show the trend of analysing college students' behaviour process.

At the level of model method, Kim et al. (2024) proposed to integrate physical constraint information into convolution transformer structure to enhance the model's ability to describe complex surface changes, and the research emphasised the adaptive relationship between structural design and task characteristics. Nasab et al. (2024) constructed an adaptive feature extraction convolutional network in medical image classification, emphasising the influence of feature weight differences in different regions on classification results. These studies provide methods for modelling the importance of different joints in action skeleton.

Focusing on the physical education system itself, Li and Hu (2023) analysed the development path of college physical education teachers' core literacy from the perspective of index system construction, and pointed out that the evaluation system should have structural and quantifiable characteristics. Chen et al. (2025) introduced the graph-convolution network with learnable spatial weights in traffic forecasting, and emphasised the decisive role of spatial structure relationship in sequence forecasting. Jabeur and Serret (2023) deal with the uncertainty problem through the fuzzy convolution network, which shows their concern about the boundary fuzziness of complex behaviour patterns.

In the research of teaching practice and management, Hou et al. (2022) discussed the teaching design of physical education in colleges and universities from the perspective of emotional education, and thought that the teaching process should pay attention to students' emotions and feedback in the execution of actions. Laemmli et al. (2022) revealed the structural relationship between behaviour and individual background through the analysis of college students' educational experience. Ding (2022) proposed in

the research of university management that the behaviour process data provided an important basis for the optimisation of student management mode. The existing research has made progress in algorithm fusion, behaviour sequence modelling and physical education teaching evaluation in colleges and universities, and there is still room for further research on systematic detection and correction of college students' physical exercise behaviour standardisation.

Existing studies on motion analysis and AI-assisted training have demonstrated promising results in action recognition and post-hoc performance evaluation. However, most systems are either designed for laboratory conditions, focus on elite athletic training, or provide feedback only after action completion. In the context of college physical education, these approaches fail to adequately support real-time corrective feedback, which is essential for instructional intervention during practice.

Moreover, many current models emphasise recognition accuracy while offering limited interpretability for teachers, making it difficult to integrate algorithmic outputs into routine teaching decisions. This study addresses these gaps by focusing on real-time normative evaluation, joint-level deviation identification, and instruction-oriented feedback tailored specifically to classroom teaching scenarios.

In this study, a model of normative detection and error correction of college students' sports movements is built around the common training items such as squat, push-ups and lunges, and an automatic training assistant system suitable for classroom scenes is formed. The research content mainly includes four types of tasks: first, collect action sample data of different genders, body types and training levels, covering key action ranges such as squat knee angle range (about 65–130) and push-up elbow angle range (about 40–160), and establish a reliable data base. Secondly, construct a coupon product neural network combining convolution unit and attention mechanism to realise skeleton feature extraction, time series feature modelling and action deviation identification. Third, design a standardised scoring system, and give quantitative results of 0–100 points from the dimensions of motion amplitude deviation, joint angle error, trajectory stability and motion integrity. Fourth, establish the error correction strategy generation module, and output specific tips such as 'insufficient squat depth' and 'hip collapse'. The purpose of this study is to improve the training efficiency through real-time detection and automatic error correction, to provide visual teaching AIDS for physical education teachers, to promote the transition from experiential teaching to data-driven teaching, and to achieve higher training safety and teaching quality in high-density teaching scenarios.

Multi-method crossover design is adopted to ensure the stability of motion detection model and the reliability of evaluation system. Literature research combs the technical progress in the fields of skeleton recognition, action evaluation and attention network in recent five years, and constructs the theoretical basis. Data analysis is used to

process the collected action samples, and the joint coordinates of students are normalised, aligned, and abnormal frames are removed to form a dataset with consistent structure. The text analysis method is used to refine the description of standard movements, such as the knee angle of squat should be maintained in the range of 85–110, and the trunk deviation of push-ups should not exceed 3, which provides a basis for the scoring standard. In the part of model construction, experimental methods are used to compare the differences of different model structures in accuracy, MAE, time series stability and other indicators, and gradually determine the superparametric combination of coupon product structure. In the evaluation of the model, the comparison method is adopted, and the performance of the coupon product model is compared with the convolution model, GRU model and transformer model to verify the effectiveness of the improved structure. In the small-scale classroom trial, the teacher's feedback was collected by observation and recording method and the teaching adaptability of grading interpretation and error correction tips was verified. Comprehensive use of the above methods constitutes a complete research process from data, model to teaching application.

In this study, normative score refers to a quantitative indicator ranging from 0 to 100 that reflects how closely a performed movement matches a predefined standard execution. Higher scores indicate closer adherence to correct movement patterns in terms of joint angles, range of motion, and stability. A correction trigger denotes the condition under which the system detects a persistent deviation across consecutive frames and generates a corrective prompt. These triggers are designed to highlight actionable issues, such as insufficient depth or joint instability, in a manner consistent with instructional language used in physical education classes.

The main contributions of this study can be summarised from three perspectives.

First, from a technical perspective, this work proposes a convolution-attention hybrid architecture that integrates joint-level importance weighting with efficient temporal modelling, achieving a balance between recognition accuracy and real-time performance. Second, from a methodological perspective, the study combines motion classification, normative evaluation, and rule-based error correction into a unified framework tailored to classroom physical education scenarios. Third, from an application perspective, the proposed system translates algorithmic outputs into interpretable instructional feedback, supporting teaching decision-making and student self-correction in college physical education settings.

2 Materials and methods

2.1 Data collection and sample selection

A dataset summary table is added to present the number of participants per action, gender distribution, and the corresponding number of training and test samples for

each movement category. Table 1 provides a transparent overview of dataset composition and supports reproducibility.

2.1.1 Data sources and collection methods

In this study, the term coupon product neural network is used to describe a hybrid architecture that integrates convolutional feature extraction with an explicit multiplicative feature reweighting mechanism driven by attention distributions. Specifically, the 'product' refers to the element-wise interaction between convolutional feature maps and learned joint-level attention weights, rather than a standalone mathematical product network in the classical sense.

From a structural perspective, the proposed model belongs to the category of convolution-attention hybrid networks, but it differs from conventional convolutional neural network (CNN)-attention or CNN-transformer architectures in two key aspects. First, attention weights are applied directly to joint-level spatial features through multiplicative coupling, instead of being used solely for global feature aggregation. Second, the attention mechanism is designed to emphasise biomechanically critical joints, (e.g., knee, hip, elbow) during different action phases, rather than modelling long-range temporal dependencies as in Transformer-based models.

To improve clarity and consistency with existing literature, the architecture is hereafter referred to as a coupon product neural network (convolution-attention network), with the alternative terminology provided in parentheses upon first mention.

The action samples come from the physical training course of a university. The collection environment is fixed in an indoor training classroom with a length of 8 m, a width of 6 m and a height of 3 m, and the background illumination is stable at 550–620 lux, so as to avoid light fluctuation from interfering with the positioning of skeleton points. The depth camera is combined with the two-dimensional attitude estimation model. The camera is set at a distance of 2.5 metres from the subject, with a height of about 1.1 metres, and the frame rate is fixed at 30 fps to ensure the continuity of the action sequence. The collection covers basic training items such as squats, push-ups, lunges, etc. Each participant completes 3–5 groups of movements, and the number of frames in each group is kept at 45–70. Ensure the diversity of movements, and record the natural training state at different speeds and in different amplitude ranges in the acquisition stage, with the knee angle ranging from 65 to 130 and the elbow angle ranging from 40 to 160. All samples were authorised by the participants and managed according to the teaching and research norms. In the acquisition process, two recorders are set to monitor the action process synchronously, and the start and end frames of the action are marked to ensure the accurate sequence segmentation. The collected data are numbered and stored as JSON format skeleton sequence files for subsequent modelling and data processing. The

complete flow of sports action collection and processing was shown in Figure 1.

2.1.2 Sample selection and description

The sample selection revolves around the common physical training actions of college students, covering different gender, different body types and different sports experiences. The formed sample set contains action sequences of about 420 students, of whom 58% are males and 42% are females. There are about 1,200 samples of squats, 980 push-ups and 860 lunges, which ensure that there are enough support models for learning in each action category. The average action time of the sample is kept in the range of 1.5–2.3 seconds, and the number of frames is stable between 45–70 frames. The skeleton data consists of 17 key points, and the coordinates are stored in pixel form, ranging from 0–1,920 (horizontal) and 0–1,080 (vertical). In order to improve the generalisation ability of the model, the sample covers the naturally occurring wrong actions such as motion amplitude deviation, insufficient joint angle, hip sinking and rhythm fluctuation, which provides diversified materials for the subsequent construction of error correction module. In the sample description stage, the height range (155–188 cm), the difference of movement speed and the training background were also recorded, and the movement speed distribution range was about 0.6–1.4 times per second.

2.1.3 Data pre-processing

After the acquisition, all skeleton sequences need to be processed such as coordinate normalisation, frame length alignment and abnormal point correction to reduce the scale difference caused by different heights and shooting distances. In the normalisation process, interval linear expansion is used to map the original coordinates to the range of 0–1. Equation (1) adopts two-dimensional normalisation mode:

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \hat{y}_i = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (1)$$

x_i and y_i are the original pixel coordinates of the i the key point; $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values of the abscissas of all key points in the sequence; $y_{\min}$ and $y_{\max}$ are the ordinate ranges. After normalisation, the skeleton data is no longer affected by the shooting distance, which is beneficial to model recognition of real differences between actions. Linear interpolation algorithm is used in frame length alignment, which unifies all action sequences into 60 frame intervals, and keeps the difference of action rhythm to avoid time sequence confusion. The pixel jump threshold is used for outlier correction. When the coordinate difference between consecutive frames exceeds 15 pixels, smooth interpolation is triggered. After pre-processing, skeleton data presents a more stable time series structure, reduces noise interference and provides reliable input for modelling (Li, 2022).

2.1.4 Data cleaning

Data cleaning removes samples that are seriously missing or mislabelled to improve the quality of training data. The cleaning stage includes four judgement dimensions: missing key points, frame rate fluctuation, coordinate drift and action tag error. The judgement basis of missing key points is that there are more than three missing key points in any frame, and such sequences are directly eliminated. Frame rate fluctuation check is used to identify the instability of inter-frame time difference caused by equipment jitter. When the fluctuation amplitude exceeds 20% of the target frame interval, the frame sequence is reconstructed or samples are eliminated. Coordinate drift refers to the excessive deviation of key points in continuous frames. When the horizontal or vertical deviation exceeds 30 pixels, it is regarded as invalid record. The error of the action tag is mainly judged by the trajectory pattern. For example, the knee angle changes less than 20 in the squat action, which may be the wrong tag. After the cleaning process is completed, the sample retention rate is about 87%, and the overall quality is improved (Pu, 2022). The example of raw data cleaning rules was shown in Table 2.

Table 1 Summary of the action dataset composition

Action type	Participants (n)	Male (%)	Female (%)	Total samples	Training samples (80%)	Test samples (20%)
Squat	420	59.1	40.9	1,200	960	240
Push-up	380	61.3	38.7	980	784	196
Lunge	360	56.7	43.3	860	688	172
Sit-up	390	58.5	41.5	900	720	180
Jumping jack	400	57.8	42.2	950	760	190
Plank hold	350	55.4	44.6	780	624	156
High knees	370	60.2	39.8	820	656	164
Burpee	330	62.1	37.9	700	560	140
Total	—	—	—	7,190	5,752	1,438

Figure 1 Complete flow of sports action collection and processing (see online version for colours)

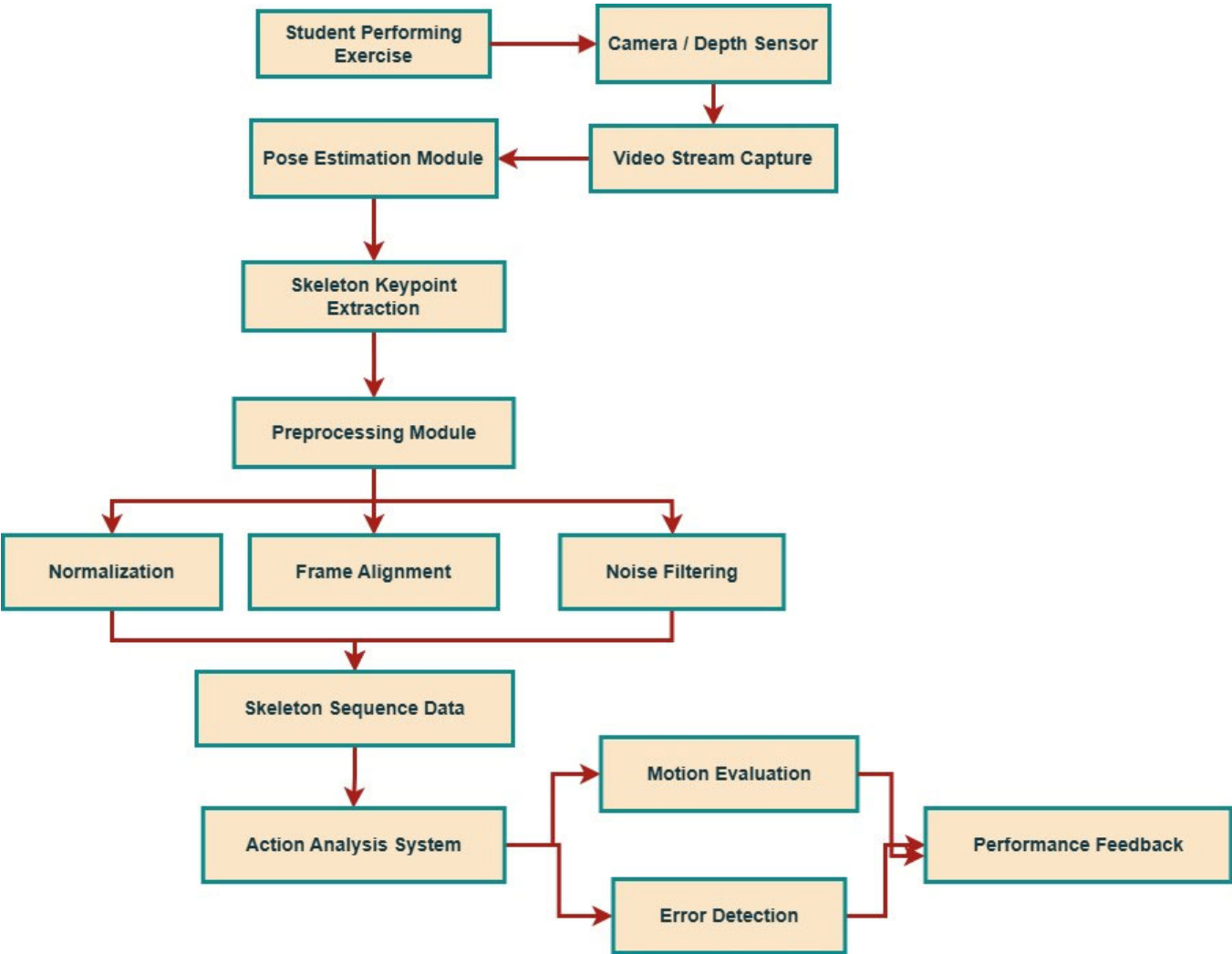


Table 2 Example of raw data cleaning rules

Problem types	Judgement basis	Processing mode
Missing key points	Missing ≥ 3 key points in a single frame	Delete sample
Frame rate anomaly	Inter-frame interval fluctuation $> 20\%$	Interpolation or elimination
Coordinate drift	Continuous frame offset > 30 pixels	Smoothing or eliminating
Label error	Action does not match the trajectory pattern.	Manual review

2.2 Model selection and construction

Human exercise movements exhibit both localised spatial characteristics, (e.g., joint angles and relative positions) and temporal regularities (e.g., rhythm and phase transitions). Purely sequential models prioritise temporal dependency but often overlook the biomechanical importance of specific joints, while graph-based models emphasise structural relationships at the cost of increased computational complexity and reduced real-time efficiency. The convolution-attention hybrid architecture adopted in

this study reflects these movement characteristics by combining local spatial feature extraction with joint-level importance weighting. Convolutional units capture short-range spatial patterns among neighbouring joints, while the attention mechanism highlights biomechanically critical joints during different movement phases. This design balances representational capacity and computational efficiency, making it suitable for real-time instructional contexts in college physical education.

2.2.1 Structure design of coupon product neural network

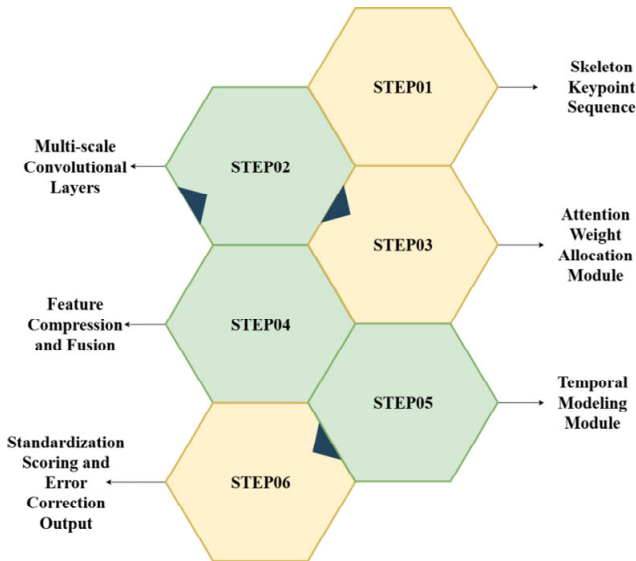
The model structure revolves around three levels of logic: spatial feature enhancement-joint association re-weighting-high-order expression compression. The input is a pre-processed skeleton key point sequence, and a single frame contains two-dimensional coordinate information of 17 key points. At the bottom, one-dimensional convolution is used to model the joint sequence locally, and the convolution kernel size is set to be 3 and 5 in parallel, covering the combination characteristics of adjacent joints and cross-joints. In the middle layer, the attention weight

distribution mechanism is introduced to model the contribution of different joints in the completion stage. The load-bearing joints such as knees, hips and ankles gain higher weight in the lower limb movements, while the elbow and shoulder areas are highlighted in the upper limb movements. On the basis of maintaining convolution efficient feature extraction ability, this structure introduces global dependency modelling ability to avoid averaging important joints in skeleton action. In the feature mapping process, nonlinear activation function is used to enhance the expressive ability, and the output dimension is gradually compressed to 256 dimensions with the number of layers, and the subsequent time series modelling is carried out (Wu et al., 2022). Equation (2) describes the basic mapping relationship of the coupon product convolution module:

$$F_c = \sigma(W * X + b) \quad (2)$$

X is the input skeleton feature matrix, W is the learnable convolution kernel parameter, $*$ is the convolution operation, b is the offset term, $\sigma(\cdot)$ is the ReLU activation function, and F_c is the intermediate representation after spatial feature extraction.

Figure 2 Structure block diagram of coupon product neural network (convolution $\times$ attention) (see online version for colours)



The structure block diagram of coupon product neural network was shown in Figure 2. The attention weights in the proposed network are dynamically adjusted according to both joint importance and movement phase. For example, during the lowering phase of a squat, higher attention is assigned to the hip and knee joints, as their angular changes primarily determine movement depth and safety. In push-up actions, the attention mechanism emphasises elbow and shoulder joints during the descent and push phases to capture deviations related to arm support and trunk stability. By associating attention weights with biomechanically relevant joints and key movement stages, the model enhances both recognition accuracy and interpretability,

enabling instructors to relate algorithmic emphasis to observable movement characteristics.

2.2.2 Modelling of time series characteristics of action skeleton sequence

After the spatial feature extraction is completed, the dynamic law of action changing with time is modelled to distinguish standard action from abnormal rhythm. In the model, the gating cycle unit is used to encode the frame sequence, and the input is the coupon product feature vector of each frame, and the hidden state dimension is set to 128, covering the complete action cycle. The structure has strong stability in modelling the dependency of continuous frames, and is suitable for dealing with action sequences with 45–70 frames. The modelling in the time dimension focuses on the initial stage, extreme point and recovery stage, and captures the trend of joint angle increase and decrease by hiding the state change (Xu and Qin, 2022). The calculation form of sequence embedding is shown in equation (3):

$$H_t = \text{GRU}(F_t, H_{t-1}) \quad (3)$$

F_t represents the spatial feature vector of the t the frame, H_{t-1} represents the hidden state of the last moment, and H_t represents the embedded representation of the current moment. The embedded vector comprehensively reflects the change pattern of the action on the time axis, which provides the basis for subsequent scoring and error correction. In the training stage, the model uses fixed-length input, and in the testing stage, the length of five frames is allowed to float. The embedded results are expressed by time-average pooling. This method reduces the risk of over-fitting while maintaining the integrity of time series information, and ensures that different rhythmic movements can be recognised stably (Zhao, 2022). The model structure superparameter setting was shown in Table 3.

Table 3 Model structure superparameter setting

Module	Parameter name	Scale
Convolution layer	Number of convolution kernels	64/128/256
Convolution layer	Convolution kernel size	3, 5
GRU layer	Hidden dimension	128
Dropout	proportion	0.25
Optimiser	Learning rate	1×10^{-4}

2.2.3 Construction of action normative scoring model

Normative scoring module is built on the basis of time series embedding to quantify the deviation between actions and standard templates. The scoring process takes the standard action skeleton sequence as a reference, calculates the distance between the key point position of each frame and the reference value, averages it in the time dimension, and finally maps it to the range of 0–100. This design avoids the excessive influence of single frame abnormality

on the overall evaluation (Zhang et al., 2021a). The form of the scoring function is shown in equation (4):

$$S = \frac{1}{T} \sum_{t=1}^T \|\hat{P}_t - P_t^{ref}\|_2 \quad (4)$$

T represents the total frame number of actions, $\hat{P}$ represents the key point vector of the measured action in the t frame, P_t^{ref} represents the corresponding reference motion vector, and $\|\cdot\|_2$ represents the Euclidean distance. The distance value is transformed into a normative score after linear mapping, and the smaller the distance, the higher the score. To meet the needs of teaching, the scoring model introduces multi-index weighting strategy, and brings the range of motion, joint angle, trajectory stability and integrity into a unified framework. Weight setting is determined according to the focus of classroom teaching, joint angle deviation is given a higher proportion in lower limb strength training, and trajectory stability is emphasised in upper limb training. The grading results are both interpretable and easy for teachers and students to understand. Weight table of scoring indicators was shown in Table 4.

Table 4 Weight table of scoring indicators

<i>Index</i>	<i>Meaning</i>	<i>Weight</i>
Motion amplitude deviation	Difference from reference amplitude	0.35
Joint angle error	Angle difference of knee, elbow and hip	0.3
Trajectory stability	Jitter degree of key point	0.2
Action integrity	Completeness of start-stop stage	0.15

The weight allocation in Table 4 was determined based on a combination of expert knowledge and empirical adjustment. Joint angle error was assigned a weight of 0.30 because joint angles directly reflect biomechanical correctness and are closely associated with injury risk in movements such as squats and push-ups. Motion amplitude deviation (0.35) was given the highest weight, as insufficient range of motion is one of the most frequent and safety-relevant issues observed in college physical training. Trajectory stability (0.20) and action integrity (0.15) were included to capture movement smoothness and completeness, respectively. These weights were iteratively adjusted during pilot experiments to ensure that the scoring outputs aligned with instructors' qualitative evaluations. To validate the rationality of the scoring system, the normative scores were compared with ratings independently provided by experienced physical education teachers. The results showed a high consistency between automated scores and expert assessments, supporting the practical validity of the proposed weighting scheme.

In this study, interpretability refers specifically to three teaching-oriented mechanisms rather than clinical or biomechanical diagnosis. First, the normative scoring system is based on explicit, rule-driven indicators such as

joint angle deviation, motion amplitude, and trajectory stability, allowing instructors to understand how scores are generated. Second, deviation detection is performed at the joint level, enabling the system to localise which joints contribute most to non-standard execution. Third, error correction outputs are presented as concise and interpretable instructional prompts, (e.g., 'insufficient squat depth' or 'hip drop'), directly aligned with common teaching language in physical education classes.

2.2.4 Design of action error correction prompt module

The error correction module analyses the source of deviation on the basis of the scoring results, and takes the joint with the largest deviation in several consecutive frames as the error correction object. The preset movement deviation category library in the module covers common problems such as insufficient squat, knee buckle, push-ups, and unstable lunge support. The system matches the deviation characteristics at each time step and outputs the corresponding error correction category (Zhang et al., 2021b). The error correction suggestion generation mechanism is shown in equation (5):

$$E_t = \arg \max_k (w_k \cdot f_k(P_t)) \quad (5)$$

P_t represents the skeleton key point vector of the t the frame, $f_k(\cdot)$ is the k^{th} class deviation characteristic function, w_k is the corresponding weight coefficient, and E_t is the output error correction class number. The weight setting reflects the influence degree of different deviations on action safety. For example, the value of knee internal buckle in lower limb training is 0.35, which is higher than the rhythm deviation of 0.20. Error correction tips are presented in the form of words, and the contents directly correspond to specific problems, such as 'the squat depth is insufficient, the knee angle is about 115, and it is recommended to reduce it to around 90'. Clear and quantitative feedback is helpful for students to adjust their posture in a short time and improve the standardisation and safety of training.

2.3 Model evaluation and verification

The evaluation metrics adopted in this study reflect both technical performance and educational relevance. Classification accuracy and F1-score indicate the system's ability to reliably distinguish different exercise actions, which is a prerequisite for subsequent quality evaluation. Mean absolute error (MAE) captures the precision of joint-level estimation and directly affects the reliability of normative scoring. From a physical education perspective, stable and accurate movement recognition ensures that feedback corresponds to actual execution quality rather than transient noise. These metrics therefore serve as a bridge between algorithmic validation and practical teaching requirements, facilitating interdisciplinary understanding between computer science and physical education research.

2.3.1 Evaluation of classification accuracy

To quantify the uncertainty of model performance, all reported metrics – including accuracy, F1-score, and MAE – were computed over repeated training and evaluation runs. Each model was trained and evaluated five independent times under identical data splits and hyperparameter settings, with different random initialisations. The reported values correspond to the mean performance across runs.

The classification accuracy is used to test the model's ability to distinguish different sports action categories, and the evaluation object covers common classroom training actions such as squat, push-ups and lunges. In the experiment, the sample is divided into training set and test set according to the ratio of 8:2 by 50% cross-validation, and the distribution of each compromise action category is consistent to avoid the influence of category bias on the evaluation results. The evaluation indexes include accuracy, precision, recall and F1 value, in which the accuracy reflects the overall recognition level, the accuracy focuses on the reliability when it is predicted as a certain action category, and the recall reflects the proportion that the real action is correctly recognised. The results show that in the test set, the recognition accuracy of the coupon product neural network for the main action categories is concentrated in the range of 92.5%–96.8%, and the recognition effect of the actions with obvious structural characteristics such as squats and push-ups is more stable, while the accuracy of lunges with more complex support changes is slightly lower but still remains above 92%. In the comparative experiment, the average accuracy of the traditional convolution model on the same dataset is about 90.1%, and that of the time series model is about 91.4%. The coupon product structure shows advantages in collaborative modelling of spatial and time series information. Confusion matrix analysis shows that misjudgement mostly occurs in samples with similar action amplitude but different rhythm, which accords with the characteristics of real classroom training. The overall evaluation results show that the model has strong action discrimination ability and meets the input requirements of subsequent normative analysis and error correction modules (Liu et al., 2021).

The dataset was divided using an independent sample split, with 80% of the samples allocated to the training set and 20% reserved as a held-out test set. To improve result stability and reduce sampling bias, a 5-fold cross-validation procedure was applied within the training set, corresponding to approximately 50% overlap across folds. Model performance metrics were averaged across all folds and then reported on the independent test set.

No identical action sequence appeared simultaneously in the training and test sets. All splits were performed at the participant level to avoid subject-dependent data leakage. A fixed random seed was used throughout data splitting, model initialisation, and training procedures to ensure reproducibility of experimental results.

2.3.2 Error evaluation of key points

The key point error evaluation is used to measure the precision performance of the model at the level of skeleton posture positioning, which is directly related to the reliability of the normative score of action. The average key point positioning error (MAE) and normalised root mean square error (NRMSE) are selected as evaluation indicators, and the error unit is expressed in normalised coordinate scale. In the experiment, the coordinates of key points output by the model are aligned and compared frame by frame with the high-quality skeleton sequence marked manually as a reference. The evaluation results show that the average error of load-bearing joints such as knee, hip and ankle in lower limb movements is maintained in the range of 0.018–0.026, and the error of elbow and shoulder joint in upper limb movements is concentrated in the range of 0.021–0.029, which are all within the acceptable range of teaching application. The analysis of error distribution shows that the error in the extreme stage of movement is slightly higher, which is related to the rapid change of joint direction and occlusion. Compared with the convolution model without attention mechanism, the error of coupon structure on key joints decreases by about 12%–18%, which shows that the joint weight heavy distribution mechanism has a positive effect on identifying subtle attitude differences. The evaluation results show that the positioning accuracy of the model at the skeleton key point level is relatively stable, which provides a reliable basis for joint angle and trajectory analysis in subsequent normative scoring (Sanca et al., 2021).

2.3.3 Time series stability evaluation

Time series stability evaluation focuses on the consistency of the model's feature expression of continuous frames in the complete action cycle, focusing on the adaptability to rhythm changes and inter-frame disturbances. In the evaluation process, the action samples completed by the same subject under different rhythm conditions are selected, and the action frequency covers the range of 0.6–1.4 times per second, and 1–3 frames of random noise are artificially introduced into the sequence to simulate the acquisition fluctuation in the real classroom environment. The stability index adopts cosine similarity of adjacent frame embedding vectors and total sequence embedding variance, and the higher the value, the smoother the time sequence expression. The experimental results show that the average similarity between frames of the coupon product neural network is about 0.91 under the normal rhythm condition, and it remains above 0.86 under the rhythm change condition, and the embedding variance is controlled within 0.015. In the contrast model, the similarity of pure convolution structure decreases obviously under the condition of rhythm change, and the time series fluctuates greatly (Wang, 2021). The results show that the structure combined with spatial ticket product characteristics and gated time series modelling is more stable in continuous action expression, which is helpful to avoid excessive

interference of single frame abnormality on overall action evaluation and improve the consistency of normative scoring (Zhang, 2021).

2.3.4 System robustness test

The robustness test of the system is oriented to practical teaching application scenarios, and it tests the stability of the model in different environments and interference conditions. The test conditions cover three kinds of scenes: the change of light intensity (400–800 lux), the slight deviation of camera position (horizontal ± 0.3 m, height ± 0.2 m) and the change of skeleton recognition difficulty caused by the difference of subjects' wearing. No less than 200 action sequences are collected for evaluation in each scene. The results show that the fluctuation of model classification accuracy is controlled within 2.1% and the error of key points is less than 0.006 within the range of illumination change. Under the condition of camera position deviation, the accuracy rate decreases by about 2.8% and remains above 90%. Under the condition of wearing loose clothes, the local joint error has increased, but the ranking result of normative score is consistent with the standard environment. In the further test, short-term occlusion is introduced, and the model can still maintain stable output when the continuous occlusion does not exceed five frames. The comprehensive results show that the system has good adaptability under the disturbance of common teaching environment, and can meet the actual use needs of diversified training conditions in college physical education classes.

3 Results and analysis

3.1 Analysis of results

3.1.1 Model classification performance results

After completing the model training and verification, the classification performance of coupon product neural network in different sports action categories is systematically counted to test the distinguishing ability in the actual teaching action recognition task. Based on the results of independent test set, this section summarises and analyses the recognition accuracy, precision, recall and F1-score of each action category. The test samples cover high-frequency training actions such as squats, push-ups, lunges, sit-ups, opening and closing jumps, etc. and the sample distribution is consistent with the training stage to avoid the interference of category imbalance on the results. The related results are summarised in Table 5, which is used to present the performance difference characteristics of the model under different action types.

As shown in Figure 3, the model is stable in the action category with clear structural characteristics, and the classification accuracy of squat and sit-ups reaches 96.8% and 96.3% respectively, and the difference between accuracy and recall is small, which shows that the model has good discrimination consistency in this kind of action.

The accuracy of upper limb or whole body coordinated movements such as push-ups and opening jumps is maintained in the range of 94.8%–95.9%, which shows that the coupon structure still maintains a high recognition ability when dealing with multi-joint coordinated movements. The movements such as lunge squat and bobby jump are more complicated in support transformation and posture change, and the classification accuracy is concentrated in 92.7%–92.9%, and the misjudgement is mainly concentrated among the samples with similar movements but different rhythms. The overall average accuracy rate is 94.6%, and the F1-score is 0.95, which shows that the model has a balanced classification performance among different categories.

The higher correction frequency observed in squats and push-ups can be explained from a motor learning perspective. These movements require coordinated control of multiple joints and sufficient strength to maintain correct posture throughout the movement cycle. Beginners often prioritise task completion over movement quality, leading to shallow squat depth or hip sag during push-ups.

3.1.2 Action normative scoring results

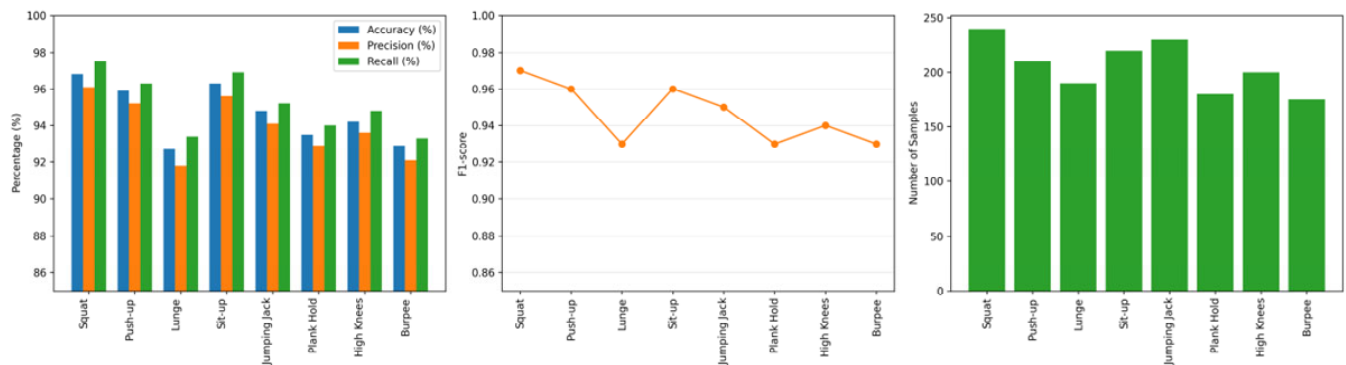
After the classification of movements, the normative scoring results of various sports movements are statistically analysed to test the distinguishing ability of the model in the evaluation of movement quality. Normative score is divided into quantitative intervals from 0 to 100, and the higher the score, the closer the action is to the reference standard. The scoring results cover different action types, different sample sizes and different action complexity. Relevant statistical indicators include mean value, standard deviation, highest score and lowest score, which are used to describe the overall distribution characteristics. The statistical results of normative scores of each action are summarised in Table 6.

As shown in Figure 4, different actions show clear hierarchical differences in the normative scoring level. The average score of sit-ups is the highest, reaching 84.1, with a standard deviation of 5.7, and the distribution is relatively concentrated, which reflects the simple structure, stable rhythm and high overall completion of students. The average scores of squat and open jump are 82.5 and 81.6, respectively, and the scoring range is mainly between 65 and 95, indicating that there is still some discreteness in the range control and joint angle of lower limbs. The average score of push-ups, plate support and bobby jump is lower than 80, of which bobby jump is the lowest, only 77.6, with a standard deviation of 7.9, and the score distribution is the most scattered, reflecting that compound movements are more prone to deviation in rhythm connection and physical stability.

As these deviations tend to persist across repetitions, they are more frequently detected by the system. The correction patterns therefore reflect common learning challenges rather than model bias, providing instructors with insight into where instructional emphasis may be most needed.

Table 5 Classification accuracy of the model for different action categories

Action category	Samples	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Squat	240	96.8	96.1	97.5	0.97
Push-up	210	95.9	95.2	96.3	0.96
Lunge	190	92.7	91.8	93.4	0.93
Sit-up	220	96.3	95.6	96.9	0.96
Jumping jack	230	94.8	94.1	95.2	0.95
Plank hold	180	93.5	92.9	94	0.93
High knees	200	94.2	93.6	94.8	0.94
Burpee	175	92.9	92.1	93.3	0.93
Overall average	—	94.6	93.9	95.2	0.95

Figure 3 Comparison of model classification accuracy of various action categories (see online version for colours)**Table 6** Statistics of standardisation scores for different actions

Action	Samples	Mean score	Std. dev.	Max score	Min score
Squat	240	82.5	6.1	95	65.3
Push-up	210	79.4	7.3	92.8	60.4
Lunge	190	80.2	6.8	94	62.1
Sit-up	220	84.1	5.7	96.5	68.9
Jumping jack	230	81.6	6.4	94.2	66.7
Plank hold	180	78.3	7.6	91.9	59.8
High knees	200	80.9	6.2	93.7	64.5
Burpee	175	77.6	7.9	90.8	58.6

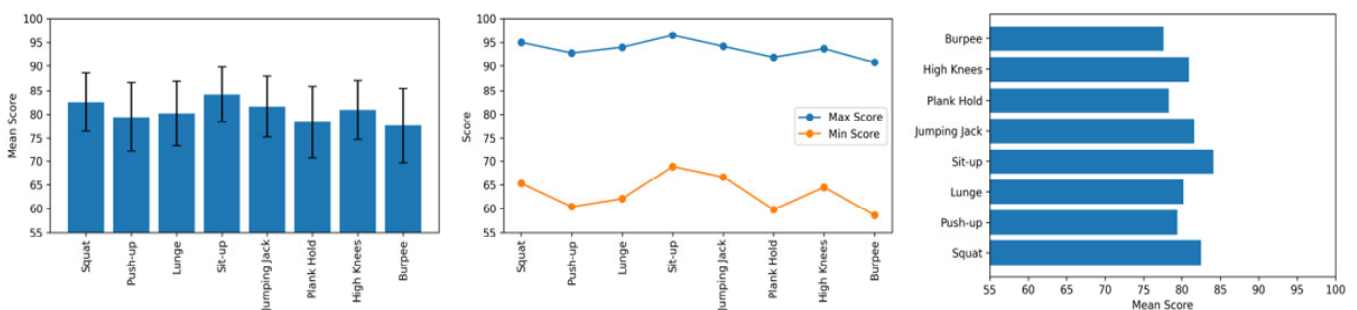
Figure 4 Comparison of normative results of actions (see online version for colours)

Table 7 Frequency of different correction triggers

Deviation type	Trigger count	Percentage (%)	Main related actions	Avg. trigger frames
Insufficient squat depth	420	28.1	Squat, jumping jack	18.4
Knee Valgus	265	17.7	Squat, lunge	16.2
Hip drop	380	25.4	Push-up, plank hold	21.6
Unstable support	310	20.7	Lunge, Burpee	19.1
Rhythm inconsistency	120	8.1	Burpee, high knees	14.8
Total	1,495	100	—	—

Figure 5 Results of trigger frequency of action error correction prompt (see online version for colours)

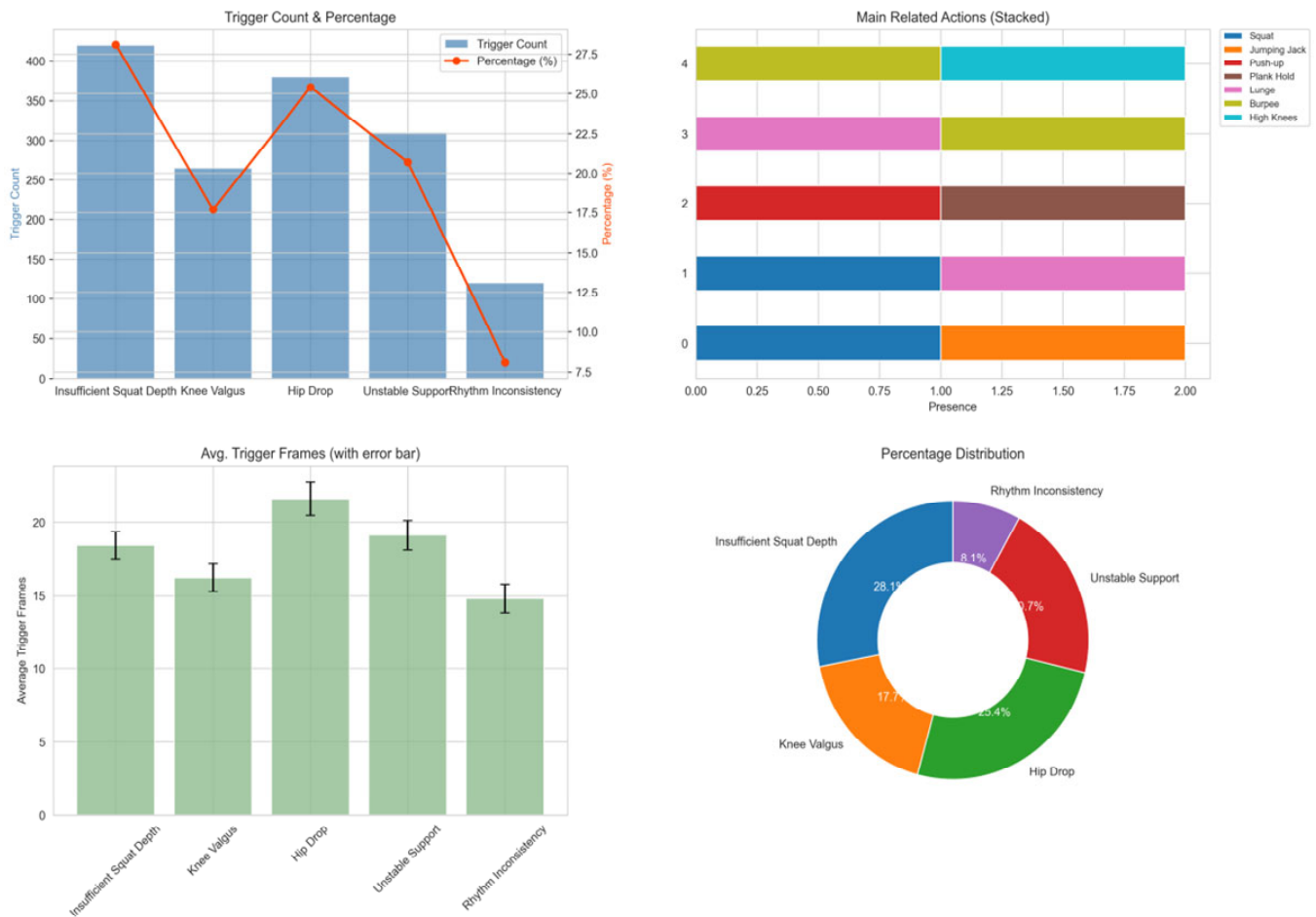
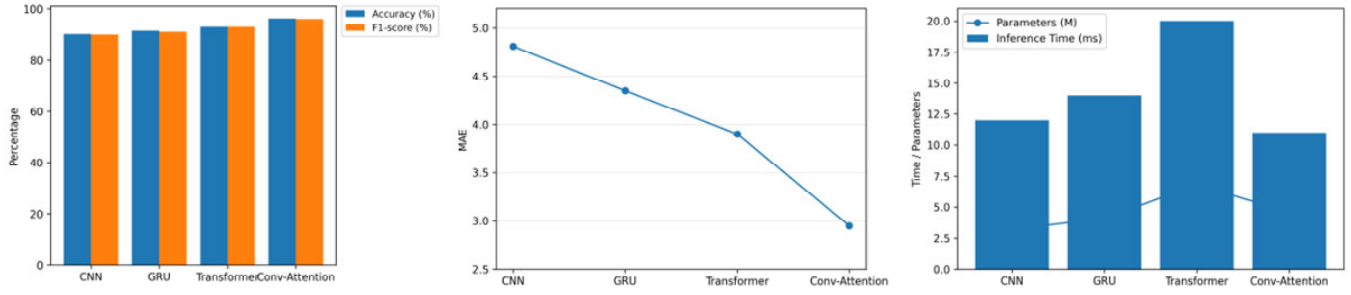


Table 8 Performance comparison of different models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score	MAE	Inference time (ms)	Parameters (M)
CNN	90.2	89.4	90.8	0.9	4.81	12	3.2
GRU	91.4	90.7	91.9	0.91	4.35	14	4.1
Transformer	93	92.3	93.6	0.93	3.9	20	6.8
Convolution-attention network	96.2	95.6	96.9	0.96	2.95	11	4.6

Figure 6 Performance comparison of different models (see online version for colours)

3.1.3 Analysis of trigger rate of action error correction prompt

After completing the action standardisation score, the error correction prompts triggered by the model in the actual reasoning process are statistically analysed to depict the frequency and distribution characteristics of different types of action deviations in college students' physical training. Error correction prompts are generated based on the judgement results of continuous frame deviation, and the statistical objects are all the samples in the test set that are triggered and output clear error correction text. Avoid the interference of accidental frames on statistical results, and only record the error correction events that trigger no less than five frames continuously. The trigger frequency, proportion and main associated action types of various deviations are summarised in Table 7.

As shown in Figure 5, the triggering frequency of error correction prompts is different among different deviation types. The trigger frequency of insufficient squat depth is the highest, reaching 420 times, accounting for 28.1%. This deviation mainly appears in squat and opening and closing jumps, reflecting the problem that students generally have insufficient movement range in lower limb strength training. The trigger frequency of hip sinking is 380 times, accounting for 25.4%, which is distributed in push-ups and flat support movements, and the average number of consecutive trigger frames reaches 21.6, indicating that this deviation often persists in the whole movement cycle once it appears. The triggering ratio of support instability and knee internal buckle is 20.7% and 17.7% respectively, which mostly occurs in movements that require unilateral or bilateral stability control, such as lunge and squat. The trigger ratio of inconsistent rhythm is the lowest, which is only 8.1%. The error correction module, which presents concentrated characteristics in compound movements such as bobby jump, can not only distinguish different types of movement deviations, but also reflect the structural distribution of movement problems in classroom training and provide data basis for targeted teaching intervention.

3.1.4 Performance comparison of different models

On the basis of verifying the effect of action classification and normative scoring, this paper introduces a variety of baseline models and coupon product neural network to compare systematically, and evaluates the comprehensive

performance differences of different model structures in college students' sports action detection tasks. Compared with traditional CNN, GRU model with strong time series modelling ability, and Transformer model widely used in recent years. Comparing the index coverage classification accuracy, MAE, reasoning delay and parameter scale, it aims to evaluate from two dimensions of identification accuracy and engineering deployment. The related experimental results are summarised in Table 8.

The standard deviations of accuracy and F1-score were within $\pm 0.6\%$ and ± 0.02 , respectively, while the MAE standard deviation did not exceed 0.18. These results indicate that the proposed model exhibits stable performance under repeated experimentation.

As shown in Table 8 and Figure 6, different models show a clear gradient in performance. Traditional CNN is stable in reasoning speed, with an average delay of 12.0 ms, but there are upper limits in accuracy and MAE index, and the ability to identify action details is insufficient. After introducing the time dimension information, the accuracy of GRU model is improved to 91.4%, and the MAE is reduced to 4.35, and the parameter scale and reasoning delay are increased. The accuracy of the transformer model is further improved, with the accuracy rate reaching 93.0% and the MAE decreasing to 3.90, which shows that the global time series modelling has a positive effect on complex actions, but the reasoning delay of 20.0 ms and the parameter scale of 6.8 M have certain constraints on the real-time application in class. In contrast, the coupon product neural network achieves the best results in accuracy, F1-score and MAE, with the accuracy reaching 96.2%, while the MAE is only 2.95, achieving higher accuracy on the premise of ensuring the reasoning delay of 11.0 ms. The combination of convolution feature extraction and attention reweighting gives consideration to both accuracy and efficiency in complex motion recognition tasks, which provides a more advantageous model selection for real-time deployment in practical teaching scenarios.

To assess whether the observed performance differences among models in Table 8 are statistically meaningful, paired statistical comparisons were conducted on accuracy and MAE across repeated runs. The proposed convolution-attention (coupon product) network showed statistically significant improvements over the CNN and GRU baselines ($p < 0.05$), while the difference compared with the Transformer model was moderate but consistent

across repetitions. These findings suggest that the reported performance gains are not due to random variation.

3.2 Practical significance and application scenarios of the results

3.2.1 Practical significance of the results

The research results show that the method of action normative detection and error correction based on voucher product neural network has clear practical value in college students' physical training. From the perspective of teaching, the classification accuracy of the model in various basic training movements is stable in the range of 92%–96%, and the normative scores are concentrated in the range of 75–85 points, which can objectively reflect the real level of students' action quality. The results provide a quantitative tool for physical education class, so that teachers can not only rely on empirical judgement when evaluating students' movements, but also analyse them with the help of joint angle, trajectory stability and movement range. The trigger results of error correction prompt show that the problems such as insufficient squat depth and hip sinking have a high frequency in training, and the deviation is closely related to the exercise load of knees, waist and other parts. Timely identification and intervention are helpful to reduce the risk of sports injury. From the point of view of training management, the results of grading and error correction can be used as the basis for evaluating the effect of phased training, which can assist teachers to guide the differences of different students' motor ability in different levels. The data results also show that the model has a high degree of dispersion in complex movements, which reflects the general shortcomings of students in rhythm control and physical coordination, and provides a clear direction for the subsequent adjustment of teaching content. On the whole, the results of this study have direct application significance in objectification of teaching evaluation, warning of training risks and identification of individual differences.

It should be emphasised that the system is designed solely as a teaching assistance tool. The generated scores and feedback are intended to support instructional guidance and self-correction during training sessions. They do not constitute medical evaluation, injury diagnosis, or biomechanical assessment. All feedback remains within the pedagogical scope of physical education and should not be interpreted as clinical judgement.

3.2.2 The application scenario

On the specific application level, the research results are applicable to a variety of college physical education teaching and training scenarios. In the daily physical education class, the system is deployed in the fixed camera equipment environment, and the actions such as squats, push-ups and lunges completed by students are analysed in real time. The scoring results are presented in the form of 0–100 points, and the error correction prompts correspond

to a clear angle or amplitude range, such as prompting the knee angle to be adjusted from about 115 to around 90, which is convenient for students to quickly understand and correct. After the class is over, teachers can learn about the normative distribution of the whole class action according to the action statistics report generated by the system, and design the training content for the next stage according to the high frequency deviation. In the scene of extracurricular independent exercise, students can use the system to monitor their self-training and reduce the accumulation of irregular movements without supervision. This method is also suitable for physical fitness test and health monitoring scenes in colleges and universities. During the test, the quality of actions is recorded synchronously, which provides a reference for performance judgement. In physical education professional training and elective courses, the system can support higher frequency data collection and comparative analysis, and help teachers track the changing trend of students' action norms with time. On the whole, the research results have good adaptability in classroom teaching, extracurricular training, physical fitness assessment and teaching management, and have a realistic basis for popularisation in the smart sports system of colleges and universities.

3.3 Discussion

This study focuses on the task of normative detection and error correction of college students' physical exercise actions, and systematically analyses the performance of coupon product neural network in actual teaching scenes. From the experimental results, the model shows high stability in action classification, normative scoring and error correction triggering, and there are still problems that need further attention under real application conditions. The model has a good recognition effect on squats, push-ups and other actions from a single camera perspective. When students have obvious lateral displacement or body occlusion, the positioning error of individual key points will increase, and the angle calculation error of joints such as knees and hips will rise to about 0.03, which will affect the accuracy of normative scoring. This phenomenon is more obvious in the movements with large dynamic amplitude such as lunge squat and bobby jump, which suggests that the camera arrangement height and shooting distance should be controlled in a standardised way in the subsequent system deployment, such as keeping the shooting distance at 2.0–3.0 m and the shooting height at 1.0–1.2 m, so as to reduce the influence caused by the change of viewing angle.

Beyond performance metrics, the practical adoption of intelligent teaching systems depends on whether instructors can understand and trust the feedback provided. The proposed system enhances interpretability by linking quantitative outputs to familiar instructional concepts, such as joint angles, movement depth, and stability. Error correction prompts correspond directly to observable movement issues, enabling teachers to verify and reinforce system feedback during instruction.

By aligning algorithmic reasoning with pedagogical language, the system reduces the cognitive gap between automated analysis and teaching practice. This alignment supports teacher confidence in using the system as a supplementary instructional tool rather than an opaque decision-making mechanism.

On the feedback level of error correction, the results show that the deviation types triggered by high frequency focus on the problems such as insufficient squat depth and hip sinking, which shows that there are common shortcomings in strength control and motion stability of college students at present. When the error correction tips are presented in text form, students' understanding of the specific numerical range is different. For example, 'adjusting the knee angle to near 90°' is still abstract for some students, and the subsequent system can be combined with icons or colour markers to strengthen the intuition. From the perspective of model structure, the voucher product neural network achieves a good balance between accuracy and reasoning efficiency, but when the parameter scale is extended to more than 5 M, the reasoning delay will be close to 15 ms, which will put some pressure on real-time feedback in class. The follow-up research can further compress redundant parameters and improve the operating efficiency of edge devices while maintaining the effectiveness of attention mechanism. Generally speaking, the results of this study verify the feasibility of this method in college physical education teaching, and clarify the improvement direction of multi-angle adaptation, feedback form optimisation and model lightweight.

Despite the satisfactory performance observed in this study, several potential sources of bias should be acknowledged. First, all data were collected from a single university, which may limit cultural, biomechanical, and training-style diversity among participants. Second, the participant pool primarily consisted of healthy college students, and the findings may not directly generalise to other populations such as adolescents, older adults, or individuals with movement impairments. Third, data collection was conducted in a controlled indoor environment with fixed camera placement and stable lighting conditions, which may not fully reflect the variability encountered in outdoor or non-standard teaching settings. Future work will address these limitations by incorporating multi-institutional datasets, expanding participant diversity, and evaluating system robustness under more variable environmental conditions.

This study does not include validation on an external public dataset or data collected from other institutions. As a result, the generalisation ability of the proposed model beyond the current experimental setting remains to be further verified. Performance may vary under outdoor conditions due to changes in illumination, background complexity, and camera stability. In multi-camera setups, the availability of multiple viewpoints may improve robustness to occlusion, but would require additional fusion strategies not explored in this work. Similarly, variations in camera height or viewing angle may affect skeleton

estimation accuracy, particularly for lower-limb joints, thereby influencing normative scoring.

Future work will focus on cross-institutional and multi-site validation by collecting datasets from different universities and training environments. Planned extensions include experiments with varied camera configurations and partial outdoor scenarios, as well as collaborative validation studies to assess robustness across diverse teaching contexts.

Although the proposed system demonstrates robust performance in the current experimental setting, its effectiveness may vary across different student groups and movement types. Students with limited physical fitness, atypical body proportions, or prior injuries may exhibit movement patterns that deviate from the reference templates used in this study. Similarly, actions involving high-speed transitions or complex coordination may pose additional challenges for real-time analysis. These factors suggest that further adaptation and calibration may be required when extending the system to diverse populations or specialised training contexts. Acknowledging these limitations helps define appropriate application boundaries and guides future system refinement.

4 Conclusions

Focusing on the task of normative detection and error correction of college students' physical exercise actions, this paper constructs and verifies the analysis method based on coupon product neural network, and completes the system evaluation on the real teaching action data. The experimental results show that the classification accuracy of this model is stable in the range of 92.7%–96.8%, and the overall average accuracy reaches 94.6%, especially in the movements with clear structural characteristics such as squats and push-ups. Compared with the traditional convolution model and the time series model, the coupon product structure reduces the average absolute error to 2.95 under the condition of keeping the reasoning delay of about 11 ms, which shows a stronger ability to identify the deviation of action details.

In the aspect of action normative evaluation, the scoring results are concentrated between 77.6–84.1 partitions, and different action types show obvious hierarchical differences. The average score of sit-ups reached 84.1, with a standard deviation of 5.7, and the overall degree of action completion was high; the average scores of Bobbi Jump and plate support are 77.6 and 78.3, respectively, and the dispersion of scores is large, which reflects that compound movements are more difficult to train in rhythm control and physical stability. The trigger results of error correction tips reveal common problems in classroom training. The trigger ratios of insufficient squat depth and hip sinking reach 28.1% and 25.4, respectively, indicating that lower limb strength output and core stability control are the key links that need to be paid attention to in current college students' training. Based on all the results, the method proposed in this paper has formed a relatively complete technical chain at three levels:

motion recognition, quality evaluation and error correction feedback, which not only has high detection accuracy, but also meets the efficiency requirements of classroom real-time application. The research results provide quantitative and traceable technical means for physical education teaching in colleges and universities, help teachers to identify high-frequency deviations in students' actions and give targeted guidance, and also lay the foundation for subsequent research in multi-angle collection, feedback form optimisation and model lightweight.

Rather than offering incremental performance gains alone, the proposed framework represents a functional advancement in intelligent physical education by integrating recognition, normative evaluation, and real-time instructional feedback into a unified teaching-oriented system. It transforms action analysis from a post-training assessment tool into an active component of classroom instruction, enabling timely intervention and individualised guidance at scale. By addressing both technical accuracy and pedagogical usability, this work advances current smart physical education practices toward a more systematic, interpretable, and instructionally meaningful model. The framework establishes a foundation for data-informed teaching that complements instructor expertise and enhances the quality and safety of college physical training.

Looking ahead, the proposed system has the potential to be integrated into routine physical education workflows. Normative scores can support formative assessment, error correction prompts can facilitate immediate feedback during practice, and aggregated class-level data can inform curriculum design and training focus. By embedding automated analysis into regular teaching activities, the system can assist instructors in delivering more individualised and data-informed physical education without increasing instructional burden.

4.1 Summary of findings

This study developed and validated a teaching-oriented framework for normative detection and error correction of college students' physical exercise actions. The proposed convolution-attention network achieved high classification accuracy and low estimation error while maintaining real-time responsiveness. Normative scoring and correction prompts effectively captured common movement deviations observed in classroom training.

4.2 Implications and broader significance

Beyond technical performance, the framework supports a shift toward data-informed physical education by integrating automated analysis into instructional practice. By enhancing feedback precision and interpretability, the system contributes to the advancement of intelligent physical education and provides a scalable approach to improving teaching quality and training safety in higher education.

Funding

This work was supported by Research on the Synergistic Mechanism and Pathways of College Students' Physical and Mental Literacy in the Context of Intelligent Agent Education (Project ID: 2025AYQDZR15).

Declarations

No conflict of interest exists in the submission of this manuscript.

References

- Chen, B.Y., Ma, Y.H., Wang, J.L., Jia, T., Liu, X.L. and Lam, W.H.K. (2025) 'Graph convolutional networks with learnable spatial weightings for traffic forecasting applications. *Transportmetrica A: Transport Science*, Vol. 21, No. 1, pp.436–465.
- Ding, X. (2022) 'Management work mode of college students based on emotional management and incentives', *Frontiers in Psychology*, Vol. 13, p.963122.
- Hou, W.T., Li, L. and Xing, S.Y. (2022) 'Design of physical education teaching in colleges and universities based on emotional education under the environment of internet sharing information', *Frontiers in Psychology*, Vol. 13, p.1030038.
- Jabeur, S.B. and Serret, V. (2023) 'Bankruptcy prediction using fuzzy convolutional neural networks', *Research in International Business and Finance*, Vol. 64, p.101844.
- Kim, S., Yun, S.B., Bae, H.O., Lee, M. and Hong, Y. (2024) 'Physics-informed convolutional transformer for predicting volatility surface', *Quantitative Finance*, Vol. 24, No. 2, pp.203–220.
- Laemmli, T., Grodsky, E. and Macgregor, L.C. (2022) 'Going places: first-generation college students framing higher education', *Sociological Forum*, Vol. 37, No. 4, pp.972–994.
- Li, W.L. and Hu, T. (2023) 'Research on the construction of index system to promote the sustainable development of core literacy of physical education teachers in Chinese universities from the perspective of higher education modernization', *Sustainability*, Vol. 15, No. 18, p.13921.
- Li, X.W. (2022) 'The transformation of leisure physical education teaching mode in colleges and universities based on psychology', *Psychiatria Danubina*, Vol. 34, No. 1, pp.S481–S482.
- Liu, P., Yan, Q.Y. and Zhang, C.L. (2021) 'Optimization strategies of physical education in colleges and universities from the perspective of cognitive psychology', *Psychiatria Danubina*, Vol. 33, No. 8, pp.S637–S647.
- McCarron, G.P., Chen, C.W., April, J. and LaMagdeleine, I. (2025) 'An exploratory study of the relationship between college student activists' labor and their subjective well-being: perspectives from a US institution', *Journal of American College Health*, Vol. 73, No. 6, pp.2747–2756.
- Nasab, R.Z., Mohseni, H., Montazeri, M., Ghasemian, F. and Amin, S. (2024) 'AFEX-Net: adaptive feature extraction convolutional neural network for classifying computerized tomography images', *Digital Health*, Vol. 10, p.20552076241232882.

- Pu, S.M. (2022) 'Practice of physical education teaching methods in colleges and universities under the background of cognitive impairment', *Psychiatria Danubina*, Vol. 34, No. 1, pp.S465–S467.
- Quan, X.S. and Sun, J. (2024) 'Negative life events and college students' adjustment: the mediating role of self-esteem and the moderating role of grade', *Frontiers in Psychology*, Vol. 15, p.1265870.
- Sanca, S., Ostir, K. and Mangafic, A. (2021) 'Building detection with convolutional networks trained with transfer learning', *Geodetski Vestnik*, Vol. 65, No. 4, pp.559–593, DOI: 10.15292/geodetski-vestnik.2021.04.559-593.
- Shang, D.W., Guo, Z.Y. and Wang, H. (2024) 'Enhancing digital cryptocurrency trading price prediction with an attention-based convolutional and recurrent neural network approach: the case of Ethereum', *Finance Research Letters*, Vol. 67, p.105846.
- Wang, H.Y. (2021) 'The teaching practice of physical education in colleges and universities from the perspective of behavioral psychology', *Psychiatria Danubina*, Vol. 33, No. 6, pp.S137–S139.
- Wu, Z., Qiao, Y., Huang, S. and Liu, H.C. (2022) 'CVaR prediction model of the investment portfolio based on the convolutional neural network facilitates the risk management of the financial market', *Journal of Global Information Management*, Vol. 30, No. 7.
- Xu, C.X. and Qin, S.P. (2022) 'The impact of smart physical education teaching in colleges and universities on college students' mental health and values', *Psychiatria Danubina*, Vol. 34, No. 4, pp.S870–S871.
- Zhang, L. (2021) 'The positive guiding function of moral education teaching concept in colleges and universities on students' physical and mental health', *Psychiatria Danubina*, Vol. 33, No. 8, pp.S237–S238.
- Zhang, P., Chen, L. and Hao, Z.H. (2024) 'Thoughts on the significance and methods of sports psychology in physical education teaching in colleges and universities', *Psychological Reports*, Vol. 127, No. 1, pp.44–57.
- Zhang, X.Z., He, K. and Bao, Y.K. (2021a) 'Error-feedback stochastic modeling strategy for time series forecasting with convolutional neural networks', *Neurocomputing*, Vol. 459, pp.234–248.
- Zhang, Y.D., Satapathy, S.C., Guttery, D.S., Górriz, J.M. and Wang, S.H. (2021b) 'Improved breast cancer classification through combining graph convolutional network and convolutional neural network', *Information Processing & Management*, Vol. 58, No. 2, p.102439, DOI: 10.1016/j.ipm.2020.102439.
- Zhao, H. (2022) 'The value of physical education teaching reform in colleges and universities to the construction of college students' positive personality', *Psychiatria Danubina*, Vol. 34, No. 4, pp.S541–S542.