scientific reports

Check for updates

OPEN

The application of suitable sports games for junior high school students based on deep learning and artificial intelligence

Xueyan Ji[1], Shamsulariffin Bin Samsudin[2]✉, Muhammad Zarif Bin Hassan[2], Noor Hamzani Farizan[3], Yubin Yuan[1] & Wang Chen[1]

In the contemporary educational environment, junior high school students' physical education is facing the challenge of improving teaching quality, strengthening students' physique, and cultivating lifelong physical habits. Traditional physical education teaching methods are limited by resources, feedback efficiency and other factors, and it is difficult to meet students' personalized learning needs. With the rapid development of artificial intelligence and deep learning technology, a new opportunity is provided for physical education innovation. This study intends to develop a Spatial Temporal-Graph Convolutional Network (ST-GCN) action detection algorithm based on the MediaPipe framework. This is achieved by integrating deep learning and artificial intelligence technologies. The algorithm aims to accurately identify the performance of junior high school students in sports activities, particularly in exercises such as sit-ups. By doing so, the study seeks to enhance the adaptability and teaching quality of physical education. Finally, this approach promotes the individualized development of students. By constructing the spatio-temporal graph model of human skeletal point sequence, accurate recognition of sit-ups can be achieved. Firstly, the algorithm obtains the data of human skeleton points through attitude estimation technology. Then it constructs a spatio-temporal graph model, which represents human skeleton points as nodes in the graph and the connectivity between nodes as edges. In HMDB51 dataset, the proposed average detection accuracy of ST-GCN action recognition algorithm based on MediaPipe framework reaches 88.3%. The proposed method has advantages in long-term prediction (> 500ms), especially at 1000ms, the values of Mean Absolute Error and Mean Per Joint Position Error are 71.1 and 1.04 respectively. They are obviously lower than those of other algorithms. ST-GCN action detection algorithm based on deep learning and artificial intelligence technology can significantly improve the accuracy of action recognition in junior middle school students' sports activities, and provide an immediate and accurate feedback mechanism for physical education teaching. This approach helps students correct their movements and enhance their sports skills. Additionally, it enables teachers to gain a deeper understanding of students' physical performance. These benefits provide strong support for the implementation of differentiated teaching.

Keywords Artificial intelligence, Deep learning, Sports game, Action recognition

Research background and motivations

As science and technology rapidly develop, artificial intelligence (AI) has penetrated all aspects of life, and the field of physical education (PE) is no exception[1–3]. In recent years, PE has received more and more attention, because it is not only related to students' physical health but also closely correlated to their mental health, teamwork ability, and the cultivation of lifelong exercise habits. However, traditional PE teaching methods often rely on teachers' intuitive teaching and students' self-perception, lacking objective and accurate data support and immediate feedback[4–6]. Especially in the process of teaching and correcting complex movements, such as sit-ups, teachers need to invest a lot of time and energy to give guidance and feedback one by one. While students may find it difficult to improve their skills quickly because of the lack of timely and accurate feedback. Therefore,

[1]Department of Sports Studies, Faculty of Educational Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia. [2]Department of Language and Humanities Education, Faculty of Educational Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia. [3]Defense Fitness Academy, National Defense University of Malaysia, Sungai besi camp, 57000 Kuala Lumpur, Malaysia. ✉email: shamariffin@upm.edu.my

how to utilize modern scientific and technological means, especially AI and deep learning (DL) technology, to optimize PE teaching methods and improve teaching effects has become one of the current research hotspots.

As a vital branch of the AI field, DL provides new ideas for PE innovation with its powerful data processing and model generalization abilities[7–9]. In the PE domain, DL technology can be applied to action recognition, posture estimation, physical fitness assessment, and other aspects to furnish teachers and students with more accurate and objective data support and immediate feedback[10,11]. The application of this technology can help students better understand and master motor skills and facilitate teachers to assess students' motor levels more accurately and make personalized teaching plans. In recent years, as an open-source cross-platform framework of Google, MediaPipe framework has shown great application potential in many fields with its high efficiency, flexibility and ease of use. MediaPipe integrates a variety of advanced computer vision and machine learning technologies, and is especially good at dealing with complex tasks such as human posture estimation and facial recognition. In the field of physical education, the introduction of MediaPipe framework provides a new solution for sports detection and evaluation, which is expected to achieve more scientific, accurate, and personalized teaching. Wei et al.[12] pointed out that the design mode of wearable AI devices came from the development of the Internet. The core of hardware is various physiological information sensors and wearable technology, and the core technology of software is wireless network transmission and statistical data processing. Huang et al.[13] realized the identification of risk factors for sleep disorders through machine learning. Huang et al.[14] further explored the relationship between the increase of vitamin B6 intake and the improvement of lung function using machine learning methods.

In junior high school, students undergo a crucial phase of physical and psychological development, making the impact of physical education (PE) particularly profound. However, due to their age characteristics and cognitive limitations, traditional PE teaching methods often struggle to engage their interest and enthusiasm. Hence, it is of both theoretical and practical importance to develop an AI-assisted sports game teaching method tailored to junior high school students. As a common physical training action, sit-ups are widely used in physical education, fitness training and rehabilitation. However, traditional sit-ups often rely on teachers' intuitive judgment and students' self-perception, so it is difficult to ensure the accuracy and effectiveness of the movements. In addition, due to the differences in students' physique and skill level, it is often difficult for teachers to give full consideration in the guidance process, which leads to some students being unable to get timely correction and feedback. This study selects sit-ups as the research focus. It aims to achieve accurate identification and evaluation of sit-ups by integrating the MediaPipe framework with the ST-GCN algorithm. The approach provides scientific data support and personalized teaching suggestions for teachers. As a result, it enhances students' training effectiveness and overall experience.

The significance of this study lies in: firstly, it provides an immediate and accurate sports feedback mechanism for junior high school students, which helps them to correct their movements and improve their sports skills. Secondly, through the data-driven personalized teaching scheme, teachers can more accurately understand the sports state of each student, implement differentiated teaching, and promote educational equity. Finally, the development of the application of sports games enriches the teaching methods of sports, stimulates students' interest in sports activities, and lays a solid foundation for cultivating lifelong sports habits.

Research statement

The significance of physical education for junior high school students is undeniable. However, traditional teaching methods no longer fully meet the demands of modern education. Therefore, exploring new approaches is essential to enhance teaching effectiveness. The integration of deep learning technology into physical education holds significant potential and promising prospects. However, current research in this area remains limited and requires further exploration and practical application. Additionally, the MediaPipe framework, a powerful machine learning tool, has been successfully implemented across various fields. Yet, its application in physical education is still relatively unexplored. Therefore, this study aims to utilize MediaPipe in sports activities to assess its feasibility and effectiveness in practical application.

Research objectives

The main objective of this study is to explore a deep learning algorithm for motor action recognition in junior middle school suitable physical education teaching under the background of artificial intelligence. (1) Describe the research background, objectives, and research necessity; (2) Summarize the research status of physical education teaching and posture estimation supported by artificial intelligence; (3) Introduce the method of artificial intelligence-assisted sports activities, and propose an algorithm of action recognition and posture estimation based on deep learning; (4) Experimental design to verify the application feasibility of the action recognition algorithm based on MediaPipe framework in real space; (5) Summarize the research contribution, limitations and future research direction.

Related works

The swift progress of AI technology has brought revolutionary changes to PE teaching. Huang et al.[15] used machine learning model to predict physiological indexes. Liao et al.[16] constructed a key point detection model for youth football training based on DL cellular neural networks. Huang et al.[17] proposed a comprehensive framework of machine learning for medical applications. It included an initial machine learning selection method, which used bootstrap simulation to calculate the confidence interval of the accuracy statistics of many models. AI can provide more scientific and precise guidance for PE teaching through data collection, analysis, and processing. Table 1 shows the research progress of action recognition technology.

Demrozi et al.[18] focused on key technologies in the field of sports AI, encompassing motion behavior recognition and motion image analysis technologies. Motion behavior recognition technology assisted athletes'

Number	Author	Theme	Main content
[18]	Demrozi et al. (2020)	Key technology of sports artificial intelligence	It focuses on sports behavior recognition technology and sports image analysis technology, and assist athletes' training and competition analysis through sensor data collection, preprocessing and motion recognition algorithm.
[19]	Agarwal & Alam (2020)	Application effect of artificial intelligence in physical education teaching	It is pointed out that artificial intelligence can monitor students' sports state in real time, provide immediate feedback and guidance, help students correct wrong actions and improve their sports skills.
[20]	Ferrar et al. (2020)	Human pose estimation based on multi-sensor fusion	In order to improve the accuracy and robustness of attitude estimation, a multi-sensor data fusion method combining inertial measurement unit, optical camera and depth sensor is proposed.
[21]	Dua et al. (2021)	Application of attention mechanism in human posture estimation	The principle of attention mechanism is introduced, and the application of spatial attention mechanism and channel attention mechanism in human posture estimation task is explained in detail.
[22]	Mazzia et al. (2022)	Estimation and recognition of human posture using bone point information	It describes a method of detecting bone points in an image through a deep learning model, and estimating and recognizing the posture based on the connection relationship between bone points.

Table 1. Research progress of motion recognition technology.

training and game analysis through sensor data collection, data preprocessing, and action recognition algorithms. Motion image analysis technology could understand the movements and tactics of athletes by analyzing game or training images. Agarwal & Alam[19] pointed out that AI could monitor students' movement status and performance in real-time and provide instant feedback and guidance to students. This helped students correct errors in movements promptly and improve their sports skills. In basketball teaching, AI could identify students' shooting actions through video analysis technology and offer suggestions on shooting posture, strength, angle, and other aspects.

Motion pose estimation is a critical research direction in the computer vision field, aiming to accurately identify human poses from images or videos. Ferrar et al.[20] proposed a human posture estimation method based on multi-sensor fusion. By fusing the data of inertial measurement unit, optical camera and depth sensor, the accuracy and robustness of attitude estimation were improved. Dua et al.[21] introduced the attention mechanism into the task of human posture estimation to improve the model's attention to key information. The principle of attention mechanism and its application in attitude estimation were introduced in detail, including spatial attention mechanism and channel attention mechanism. Mazzia et al.[22] used the information of bone points to estimate and identify human posture. Firstly, the bone points in the image were detected by the deep learning model, and then the attitude estimation was carried out based on the connection relationship among the bone points.

It is worth noticing that the application of MediaPipe framework in different fields also provides valuable reference for this study. For example, in the field of medical rehabilitation, MediaPipe has been used for the evaluation and training of patients' motor ability, and provides accurate data support for rehabilitation teachers by monitoring patients' motor posture in real time. These successful cases show that the MediaPipe framework has strong application potential and wide adaptability.

Proposed solution

AI-assisted sports activities in junior high school

From data analysis and motion pose estimation to virtual reality training, smart wearable devices, and intelligent referee systems, AI technology has brought revolutionary changes to various aspects of sports activities. In sports activities, data analysis is crucial[23,24]. AI technology collects and analyzes a large amount of match data, athlete training data, etc., to provide coaches and athletes with precise data support and decision-making basis. Specifically, AI can employ machine learning algorithms to mine athletes' historical data, discover patterns and trends, and develop personalized training plans and game strategies for athletes[25–27].

Computer vision technology is a vital tool for motion pose estimation and motion behavior recognition[28,29]. By capturing students' motion images with cameras and utilizing DL algorithms to process and analyze the images, it is possible to accurately identify students' motion poses and behaviors. This technology can furnish students with real-time feedback and guidance to help them correct errors in movements and improve their sports skills.

DL technology is one of the core technologies in the AI field, which automatically learns feature representations in data and achieves high-precision prediction and classification[30–32]. In PE teaching, DL technology can be applied to motion pose estimation, physical fitness assessment, motion behavior recognition, and many other aspects. By training neural network models, precise analysis and processing of students' motion data can be achieved, offering scientific decision support for teachers.

Sensor technology can monitor students' motion data and physiological indicators in real-time, such as heart rate, number of steps, movement trajectory, etc. By transmitting these data to smart devices or cloud servers for analysis and processing, personalized exercise suggestions and guidance can be provided to students. Meanwhile, sensor technology can also be used to monitor students' motion status and safety conditions, ensuring their safety during sports activities. The application of AI technology in sports activities is summarized, as detailed in Table 2.

Human pose estimation method and the Mediapipe framework in sports games

Human pose estimation refers to the extraction of the positions of human joints from images or videos using computer vision technology, thereby estimating and recognizing human body poses[33–35]. In the field of sports, human pose estimation is mainly used for athletes' technical analysis, motion correction, and development

Technical name	Application field	Major function
Computer vision technology	Motion pose estimation	Identify student motion pose and provide immediate feedback
	Motor behavior recognition	Identify students' motor behavior and offer training suggestions
DL technology	Physical fitness assessment	Analyze the students' sports data and assess their physical fitness level
	Decision support	Provide scientific decision support for teachers
Sensor technology	Motion data monitoring	Real-time monitoring of students' motion data and physiological indicators
	Safety monitoring	Monitor students' motion status and safety conditions

Table 2. The application of AI technology in sports activities.

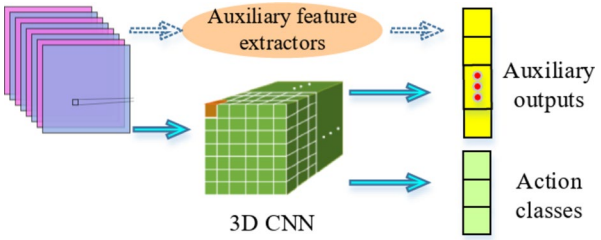

Fig. 1. Feature extraction of 3D CNN.

of training plans. Depending on the task, human pose estimation can be divided into two-dimensional (2D) and three-dimensional (3D) pose estimation. Based on the recognition of individuals, it can be classified into single-person and multi-person pose estimation. The studies involving human participants were reviewed and approved by Department of Sports Studies, Facility of Educational Studies, Universiti Putra Malaysia Ethics Committee (Approval Number: 2022.200219819996). The participants provided their written informed consent to participate in this study. All methods were performed in accordance with relevant guidelines and regulations.

The 2D human pose estimation method is mainly employed to extract the 2D coordinate information of human joints from images or videos. This method is primarily used for technical analysis and motion correction of athletes. Common 2D human pose estimation methods include template matching-based and DL-based methods.

The template matching-based method is a traditional approach to human pose estimation. This method first establishes a template library of human poses and then matches the input image with the templates in the library to find the most similar template, thus obtaining the position information of human joints[36,37]. This method is simple and easy to implement, but it has poor capabilities in handling intricate backgrounds and occlusions.

The DL-based method is currently the mainstream approach in the 2D human pose estimation domain. This method trains the DL model to automatically extract the position information of human joints from images or videos, involving key point positions, joint angles, and body poses. Moreover, this method has strong feature extraction capabilities and generalization abilities, enabling it to deal with complex occlusions and backgrounds.

3D CNN achieves direct processing of 3D data by introducing 3D convolution kernels in the convolutional layers[38–40]. These kernels slide in the 3D space (height, width, and depth) and perform convolution operations with local regions in the input data to extract 3D features. This operation enables 3D CNN to capture spatial relationships and temporal dynamics in 3D data. In 3D CNN, components such as activation functions, pooling, fully connected (FC), and convolutional layers appear alternately, collectively forming the entire network structure. Convolutional layers extract features through convolution operations; pooling layers are utilized to mitigate the size of feature maps while retaining critical information. FC layers merge features and output the final classification or regression results, and activation functions introduce nonlinearity to enhance the network's representation ability. By connecting a series of auxiliary output nodes to the last hidden layer of CNN, the extracted features during the training process more accurately approximate the computed high-level behavioral motion feature vectors, as revealed in Fig. 1.

MediaPipe is an open-source framework developed by Google for implementing machine learning models in the media processing domain, encompassing images, videos, and audio, among others. This study selects sit-ups as the research focus, aiming to achieve accurate identification and evaluation of sit-ups by integrating the MediaPipe framework with the ST-GCN algorithm. The approach provides scientific data support and personalized teaching suggestions for teachers. As a result, it enhances students' training effectiveness and overall experience. The core framework of MediaPipe is implemented in C++ and supports languages like Java and Objective C. Its main concepts include packets, streams, calculators, graphs, and subgraphs. MediaPipe graphs are directed, with packets flowing from data sources into the graph until they leave at the sink node.

Taking a hand key point detection model as an example, as presented in Fig. 2, the model operates on the cropped image area defined by the hand detector and returns high-fidelity 3D hand key points.

BlazePose is a pose estimation algorithm based on the MediaPipe framework, specifically designed for real-time human pose estimation. It features high accuracy, efficiency, and real-time performance, suitable

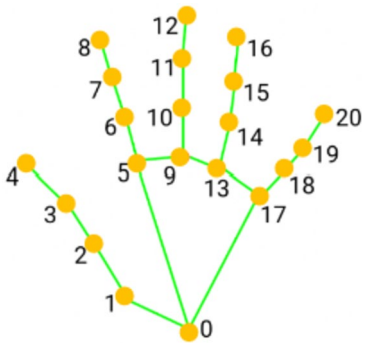

Fig. 2. Key point location of 3D hand joint coordinates based on MediaPipe.

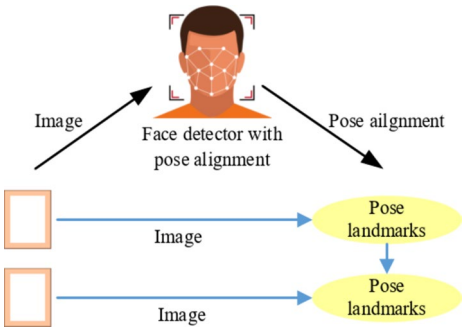

Fig. 3. Human pose prediction process based on BlazePose algorithm.

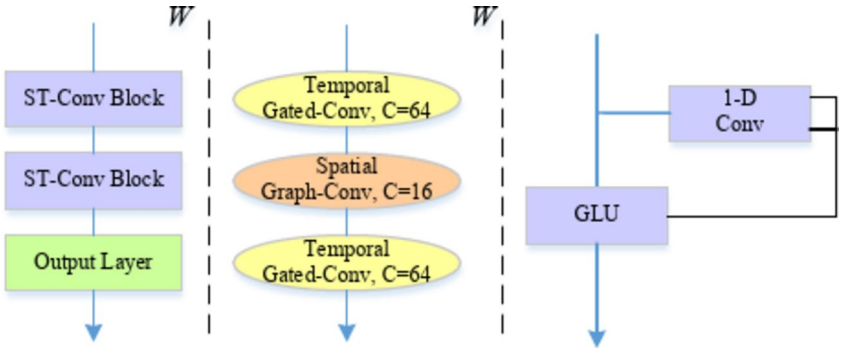

Fig. 4. ST-GCN architecture.

for various platforms including mobile devices, desktop computers, and embedded systems. BlazePose uses a lightweight neural network architecture based on MobileNetV3 design, with lower computational and parameter requirements. This network structure achieves efficient feature extraction and information fusion through depth-wise separable convolutions and lightweight attention mechanisms. The BlazePose algorithm-based human pose prediction process is illustrated in Fig. 3.

Fine-grained action segmentation and recognition based on DL

In the domain of video analysis and understanding, fine-grained action segmentation and recognition are pivotal research directions. Traditional action recognition methods typically concentrate on class labels for entire videos or video segments, while fine-grained action segmentation requires precisely segmenting each action in the video and assigning corresponding class labels[40–42]. In the fine-grained action segmentation and recognition of human skeletal motion, this study applies the Spatial Temporal-Graph Convolutional Network (ST-GCN) method (Fig. 4).

The pseudo code of ST-GCN is shown in Table 3.

In each time step, the skeleton diagram structure is rolled up in space to capture the spatial relationship between key points[41–43]. This is usually achieved by defining an adjacency matrix, which describes the connection

Input:
X: Input spatio-temporal graph data with dimensions (N, C, T, V, M), where
N is the number of samples
C is the number of feature channels (e.g., joint coordinates dimensions)
T is the number of temporal frames
V is the number of joints
M is usually 1, indicating no additional dimension (could be more for e.g., joint types)
A: Graph structural information, i.e., adjacency matrix with dimensions (V, V)
Θ: Network parameters
Output:
Y: Classification results for the actions
Procedure:
1. Initialize network parameters Θ
2. For each sample x ∈ X:
a. Construct spatio-temporal graph: Based on A and the temporal dimension T of x, build the complete spatio-temporal graph
b. Initial feature representation: Typically, x is considered as the initial feature map
3. Repeat the following steps for L iterations (L is the number of ST-GCN blocks):
a. For each ST-GCN block:
i. Spatial graph convolution:
- For each temporal frame t ∈ [1, T], apply graph convolution operation to transform the feature map spatially
- Use A as the adjacency matrix to aggregate features based on the connectivity between nodes
ii. Temporal convolution:
- Apply standard temporal convolution (e.g., 1D convolution) to the spatially transformed feature map to capture temporal dependencies
iii. Batch normalization and nonlinear activation:
- Apply batch normalization to the output of temporal convolution
- Use a nonlinear activation function (e.g., ReLU)
b. (Optional) Residual connection:
- Add the input of the ST-GCN block to its output to facilitate gradient flow
4. Apply global average pooling:
- Perform average pooling over the temporal and joint dimensions of the output from the last ST-GCN block
5. Classification layer:
- Use a fully connected layer to map the pooled features to the class space
- Apply softmax function to obtain classification probabilities Y
6. Return classification results Y

Table 3. The pseudo code of ST-GCN.

relationship between key points. Between adjacent time steps, the time graph convolution operation is carried out on the skeleton graph structure to capture the position change of key points. This can be achieved by treating the key points of adjacent time steps as nodes in the graph and defining edges between these nodes.

The internal connections of joints within a single frame are represented by the adjacency matrix A and the identity matrix I representing the self-connections. In the case of a single frame, ST-GCN using a partitioning strategy can be achieved through the following equations:

$$f_{out} = \Lambda^{-\frac{1}{2}}(A+I)^{-\frac{1}{2}} f_{in} W \tag{1}$$

$$\Lambda^{ii} = \sum\nolimits_j \left(A^{ij} + I^{ij}\right) \tag{2}$$

W indicates that weight vectors of multiple output channels are superimposed to form a weight matrix.

In the spatiotemporal context, input feature maps can be represented as (C, V, T) dimensional tensors. Graph convolution is achieved by performing$1 \times \Gamma$ standard 2D convolution and multiplying the resulting tensor with the normalized adjacency matrix$\Lambda^{-\frac{1}{2}}(A+I)\Lambda^{-\frac{1}{2}}$ in the second dimension.

In the distance partitioning strategy,$A_0 = I$and$A_1 = A$, there are:

$$f_{out} = \sum\nolimits_j \Lambda_j^{-\frac{1}{2}} (A_j) \Lambda_j^{-\frac{1}{2}} f_{in} W_j \tag{3}$$

Similarly, it can be obtained that:

$$\Lambda_j^{ii} = \sum\nolimits_k \left(A_j^{ik}\right) + \alpha \tag{4}$$

For every adjacency matrix, there is a learnable weight matrix M. That is, the equation before the substitution is:

$$f_{out} = \Lambda^{-\frac{1}{2}} [(A+I) \otimes M] \Lambda^{-\frac{1}{2}} f_{in} W \tag{5}$$

represents the product of elements between two matrices. The mask M is initialized as a matrix with all 1.

This study uses sit-up exercises in sports activities as an example. In sit-up action detection and counting, the analysis and processing of action data often involve iterative convergence and nonlinear fitting methods. The characteristic of sit-up actions is the periodic variation of the body between lying down and sitting up, which can be detected by analyzing the positional changes of a certain body part, such as the head or hips. To ensure the normativity and accuracy of sit-up actions and prevent excessive involvement of other body parts, this study constructs a reasonable penalty function by setting loss values and corresponding penalty factors for each joint or body part. The calculation of the penalty value can be written as:

$$S = \sum_{s_i}^{e_i} (\alpha \cdot loss_{wriet} + \beta \cdot loss_{knee} + \gamma \cdot loss_{ankle}) \tag{6}$$

α,β, and γ are all penalty factors; $loss_{wriet}$, $loss_{knee}$, and $loss_{ankle}$, represent the loss values of the wrist, knee, and ankle, respectively. The calculation for the loss value of each part of the body can be expressed as:

$$loss_j = \sqrt{(\alpha_j - threshold)^2} \tag{7}$$

α_j refers to the parameter value calculated for each part of the current body.

The thresholds and penalty factors for parameters of various body parts are outlined in Table 4.

In order to ensure that ST-GCN model can accurately and robustly deal with all kinds of complex and bad data conditions, this study has carried out detailed design and implementation in data preprocessing and enhancement. Firstly, key frames are extracted from the input video at a fixed frame rate to ensure that the model can handle consistent data input. Then, the extracted video frames are normalized, including image size adjustment, color space conversion (such as from RGB to gray image or YUV space) and pixel value normalization to improve the adaptability of the model to different lighting and color conditions. The pre-trained human detection model is used to detect the human body in the video frame to determine the position and size of the human body. According to the detection results, the video frame is cut, and only the area containing human body is reserved to reduce the interference of irrelevant information. In the aspect of data enhancement, the motion blur effect is simulated. By smoothing the video frame on the time axis or adding a blur filter, the image quality is simulated when the camera is moving rapidly or shaking.

Evaluation

Datasets collection

The experimental data of this study include the HMDB51 dataset and video data from 5 experimental subjects. HMDB51 is a widely used action recognition dataset that contains complex actions from various movie clips. These videos typically contain various interfering factors such as complex backgrounds, lighting changes, and camera movements due to their origin from movie scenes. In this study, video data from five subjects are used, and these subjects are tested in five different environments to evaluate the detection accuracy of the ST-GCN algorithm proposed here in different scenarios. All the data involved in the study comply with data protection and privacy laws. All personal information is anonymized during data processing to ensure the privacy and data security of participants. This includes strict control over the storage, processing and use of data to prevent unauthorized access and disclosure. The research methods and procedures all follow the research ethics and operational norms. During the experiment, the safety of the experimental environment, the physical health and psychological safety of the participants are ensured. In addition, the technical specifications of data collection, recording and analysis are followed, which ensures the scientific research and the reliability of data. It is ensured that all participants have provided written consent before participating in this study and have been approved by the ethics committee of their school. Before the experiment began, the research purpose, methods, potential risks, and participants' rights were explained in detail to ensure their full understanding and voluntary participation. All data collection and processing complied with ethical standards, and participant privacy was protected. Informed consent forms, containing detailed terms, were distributed and required to be signed by the participants' parents before they could continue. Each subject completed a certain number of sit-ups under the prescribed exercise standards, and recorded their sports videos through cameras installed in the experimental environment. Video data is used to test the performance of the algorithm in a relatively controlled environment to ensure the robustness and reliability of the algorithm. The video data are recorded by 5 experimental subjects in five different environments and used to test sit-up action detection in relatively controlled environments, and video samples.

Regarding data privacy and security, this study implements rigorous measures to protect participants' identities and personal information. During data collection and storage, all personal identifiers undergo anonymization procedures. Written informed consent is obtained from both participants and their parents, with detailed explanations of the research objectives, methodologies, and potential risks. All data transmissions and storage processes employ encryption protocols to prevent unauthorized access and misuse. Regular data backup procedures and recovery tests are conducted to ensure data integrity and availability.

Although the proposed methodology relies on high-quality input data, such data may be affected by environmental factors including lighting conditions and camera angles. To mitigate these influences, this study adopts multiple countermeasures. Specifically, data normalization and augmentation techniques are applied, encompassing image resizing, color space conversion (e.g., from RGB to grayscale or YUV space), and pixel

Parts of body	Penalty thresholds	Penalty factors
Wrist	35°	0.5
Knee	15°	0.2
Ankle	1.25	0.4

Table 4. The thresholds and penalty factors for parameters of various body parts.

value normalization to enhance model robustness against illumination variations. Furthermore, the training dataset incorporates data collected from diverse camera angles and environments to strengthen the model's generalization ability across different scenarios.

Experimental environment

Hardware configuration: CPU: Intel(R) Core(TM) i7-8700; GPU: NVIDIA T4. Software configuration: development environment: python 3.7.1 + CUDA 10.1, image processing: cv2 4.6.0.66, mediapipe 0.8.10.

To reduce the operational complexity of the system, this study has developed a user-friendly graphical user interface (GUI). The GUI simplifies steps such as model configuration, data input, and result output, making the system accessible to teachers and students without technical backgrounds. It offers a one-click initialization feature that automatically sets model parameters, eliminating the need for manual adjustments. An intuitive model selection interface allows users to choose pre-trained models like ST-GCN or other suitable models via drop-down menus or icons. Additionally, comprehensive user manuals and instructional videos are provided to guide users through system setup and usage. To further enhance user experience, a suite of automation tools has been integrated, including data preprocessing and model training tools. These tools automate tedious tasks such as data cleaning, format conversion, and model parameter tuning, thereby reducing the operational burden on users.

Parameters setting

The initial learning rate of the ST-GCN action recognition algorithm is set to 0.001. The number of iterations is set to 100 to ensure that the model is sufficiently trained but not overfitted; weight decay is 0.0001; the batch size is 32 to ensure the stability and speed of model training. Adam optimizer is chosen to automatically adjust the learning rate for each parameter to accelerate model convergence and reduce oscillation. Dropout layers are added to certain layers of the model to prevent overfitting.

Before entering the processing pipeline, all collected video data underwent rigorous anonymization. Specifically, during recording, no personally identifiable information (such as facial features) is captured. When storing the data, unique identifiers are used instead of real names or student numbers. To prevent data leakage, all data in transit, including that uploaded to the cloud, are encrypted using industry-standard protocols. Additionally, data stored on the server side are further secured with robust encryption algorithms. To ensure all operations complied with ethical standards and had necessary legal backing, this study establishes detailed data access policies and ensured transparency throughout their implementation.

Performance evaluation

In the HMDB51 dataset, the proposed algorithm's average detection accuracy reaches 88.3%. In the video dataset from 5 different scenarios, the proposed algorithm's average detection and counting accuracy of sit-up movements are depicted in Fig. 5. It can be observed that in diverse testing scenarios, due to the complexity of the environmental background, the accuracy of the action recognition algorithm is indeed affected and fluctuates to a certain extent. When the background contains many moving objects, textures, or colors similar to the target action, the algorithm may misidentify these background elements as part of the target, leading to a decrease in accuracy.

Furthermore, the proposed algorithm is compared with other mainstream action recognition algorithm models on the HMDB51 dataset to verify the proposed method's effectiveness. Among them, Spatial-Temporal Synchronous Graph Convolutional Networks (STS-GCN) can directly capture local temporal-spatial correlation at the same time. It constructs a local spatio-temporal graph, connects the single spatial maps with adjacent time steps into a map, and captures the complex local spatio-temporal correlation in these local spatio-temporal graphs through the convolution module of spatio-temporal synchronization maps. MotionMoxer is an algorithm related to motion generation, motion simulation, or similar fields. Figures 6 and 7 respectively show the comparison results of different algorithms in Mean Absolute Error (MAE) and Mean Per Joint Position Error (MPJPE) indexes. The results reveal that the proposed method has strengths in long-term prediction, such as over 500ms, especially at 1000ms, MAE and MPJPE values are 71.1 and 1.04, respectively, which are markedly lower than other algorithms.

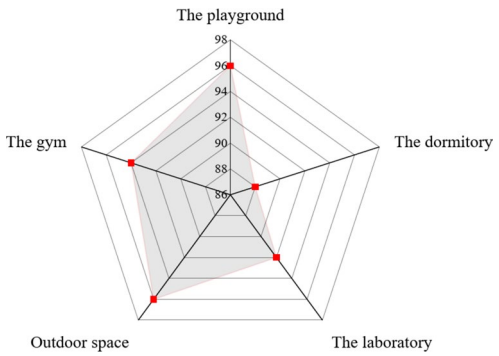

Fig. 5. Detection and counting accuracy in diverse scenarios.

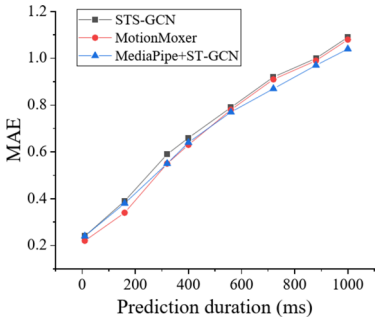

Fig. 6. Comparison of different algorithms on the MAE index.

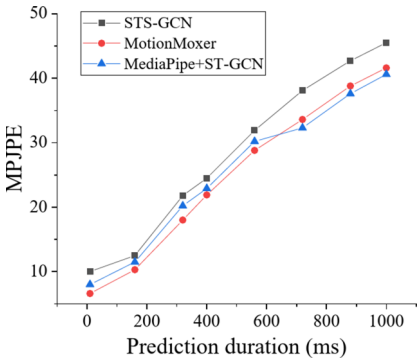

Fig. 7. Comparison of various algorithms on the MPJPE index.

Discussion

In PE, AI can provide more accurate data support for coaches and offer more personalized and appropriate sports games for junior high school students. With the swift progress of DL technology, the algorithm based on ST-GCN has garnered prominent outcomes in the human action recognition domain[44,45]. During model deployment, infrastructure and equipment requirements are crucial. High-performance computing resources and storage devices are essential for stable model operation and efficient data processing. However, in practical applications, such as sports teaching, hardware resources are often limited, and costs are constrained. On one hand, high-performance GPUs and CPUs are vital for model training and inference. These devices are typically expensive and energy-intensive, making it difficult for many schools and educational institutions to afford the long-term operational costs. On the other hand, large amounts of video data require substantial storage space, and high-performance storage devices are costly and require regular maintenance and updates. These infrastructure and equipment requirements not only increase the cost of model deployment but also limit the model's widespread adoption and application. In particular, in remote areas or resource-poor schools, hardware limitations may prevent full utilization of the model's advantages to aid teaching. To mitigate the impact of infrastructure and equipment requirements on model deployment, cloud computing can provide flexible computing resources and storage space, dynamically allocated based on actual needs. Deploying the model on cloud servers leverages the cloud's powerful computing capabilities to accelerate training and inference, reducing local hardware demands. Edge computing pushes computing tasks and data storage to the network edge, reducing data transmission latency and improving processing efficiency. By deploying the model on edge devices near the data source, such as smart cameras or smartphones, real-time data processing and feedback can be achieved, reducing bandwidth and latency. This technology is especially suitable for sports teaching scenarios with high real-time requirements, providing students with more timely and accurate guidance.

The ST-GCN motion detection algorithm based on the MediaPipe framework constructed in this study realizes accurate recognition of sit-ups by constructing a spatiotemporal graph model of the human skeletal point sequence. Initially, the algorithm obtains the data of human skeletal points through pose estimation technology. Subsequently, a spatio-temporal graph model is constructed to represent human skeletal points as nodes in the graph, and the connection relationship between nodes is represented as edges. Zhang & Tao[46] found that there were two types of edges in spatiotemporal graphs. Spatial edges represent the connection between nodes at the same time step, while temporal edges indicate the connection between the same nodes at different time steps, which is consistent with the findings of this study. Shu et al.[47] pointed out that through multiple layers of spatiotemporal graph convolution operations, the ST-GCN algorithm can capture spatial and temporal changes in human skeletal points, thereby achieving accurate recognition of sit-up actions. Table 5 shows the advantages of this study compared with the research in this field.

Paper/research	Framework/method	Spatio-temporal graph model construction
Zhang & Tao (2020)	ST-GCN	Mention Spatio-temporal graph, the specific implementation is unknown.
Shu et al. (2020)	ST-GCN	Multi-layer Spatio-temporal graph convolution
This study	MediaPipe + ST-GCN	Detailed Spatio-temporal graph model (space edge, time edge)

Table 5. Research progress of motion recognition technology.

Leveraging personalized teaching, precise data analysis, and gamified teaching, more high-quality and efficient PE services can be furnished to students. The ST-GCN model effectively integrates the spatial position and temporal dynamic information of human skeleton points by constructing a spatio-temporal graph model and realizes high-precision recognition of complex motion patterns. This technical breakthrough provides a powerful tool for action analysis in physical education teaching. Real-time data acquisition and processing based on attitude estimation technology enables the system to provide students with action feedback immediately, help them adjust their posture in time, and improve the training effect. This instant feedback mechanism is particularly important for beginners, which can significantly shorten the time to master skills.

When using video data for pose recognition and motion analysis, it is crucial to prevent the leakage of students' facial features and identity information. To this end, this study employs techniques such as facial blurring and implements strict encryption measures during data storage and transmission. However, as technology continues to evolve, privacy protection policies must also be updated and refined to address potential new risks of privacy breaches. When storing and analyzing large amounts of student data, it is essential to establish strict data access and control mechanisms to prevent unauthorized access and misuse. Regular data backups and recovery tests are also necessary to ensure data integrity and availability.

In terms of the significance for PE practice, this study integrates AI and DL technologies into physical activities, providing new insights for PE in junior high school and significantly enhancing teaching and learning experiences. Real-time feedback and personalized guidance based on precise motion detection effectively improve students' motor skills and engagement. This study also highlights the potential of technology in supporting differentiated teaching strategies, offering tailored instructional support based on the diverse needs and abilities of individual students. However, the successful application of these technologies in practice requires careful consideration of infrastructure, teacher training, and student privacy. Future research should explore the practical challenges and opportunities of deploying these technologies in diverse educational settings to maximize their impact on PE outcomes.

This study also has several limitations. First, the small sample size may affect the generalizability of the findings. The study focuses on a limited number of participants and examines only one type of exercise—sit-ups. While this narrow scope facilitates in-depth analysis, it restricts the applicability of the results to other types of physical activities. Future studies should expand the sample size, include more diverse participants, and extend the study to a wider range of exercises to enhance the robustness and generalizability of the results. Moreover, despite extensive efforts in data preprocessing and augmentation to enhance the model's robustness against poor data conditions, the model's generalization ability remains somewhat limited. Particularly when dealing with rare or complex sports activities, the model's recognition performance may decline. This is mainly due to the limitations of the training dataset, which prevents the model from fully learning the characteristics of these activities. To improve the model's generalization, future research could explore a broader range of sports activities, including various sports, dance, and fitness classes. By collecting video data of these activities and constructing a more diverse training dataset, a more robust and generalizable model can be trained. For the recognition of fine-grained or rapidly changing motions, future research could investigate more advanced deep learning algorithms and techniques, such as attention mechanisms and memory networks. These techniques can help the model better capture and understand motion details, thereby improving recognition accuracy and robustness. The proposed system in this study exhibits a primary limitation in its dependence on advanced hardware and computational resources, which may hinder adoption by financially constrained schools. To address this issue, cloud computing and edge computing technologies present viable solutions. Cloud computing provides flexible computational resources and storage capacity, enabling schools to access the model without requiring local high-performance hardware. Edge computing, conversely, processes data near its source, reducing transmission latency and improving processing efficiency - particularly advantageous for PE scenarios demanding real-time responsiveness. An additional challenge lies in the substantial processing power required for real-time analysis and feedback, potentially causing latency issues. To minimize delays, this study implements model efficiency optimizations including lightweight neural network architectures and optimized convolution operations. Furthermore, deploying the model on high-performance hardware significantly reduces processing time and latency. Future work may explore additional optimization strategies such as model pruning and quantization to enhance real-time performance.

Conclusion, and future works

Research contribution

The development of AI drives the continuous innovation of sports activities, and the action recognition and attitude estimation methods in sports activities are constantly updated. In this paper, a MediaPipe framework-based ST-GCN method is proposed for the fine-grained segmentation and recognition of human skeletons. Between adjacent time steps, a time graph convolution operation is performed on the skeleton structure to

capture the position changes of key points. The research results facilitate understanding the information in motion scenes more comprehensively and improve the recognition accuracy of specific movements.

Future works and research limitations

Although this study achieves good performance on a specific dataset, the model's generalization ability remains a challenge. This means that when the model is applied to new, unseen data, its performance can deteriorate. In future studies, larger and more diverse datasets should be collected to cover a wider range of motor movements and changes in detail. This helps to foster the model's generalization ability and recognition accuracy. During data collection, the diversity and complexity of sports teaching activities often result in highly heterogeneous video data. Differences in teaching styles and movement performances across various classes, teachers, and students make data labeling and preprocessing particularly challenging. Additionally, in some movements, students may exhibit varied forms due to insufficient strength or lack of skill proficiency, making it difficult for the model to accurately capture and recognize these actions. In future research, efforts should be made to ensure the diversity and representativeness of the data while employing data augmentation techniques to enhance the model's generalization ability. Regarding the variability in student performance, the training dataset can be enriched by introducing more student samples and types of movements, thereby improving the model's recognition accuracy. While the proposed method demonstrates excellent performance in recognizing structured movements like sit-ups, its effectiveness in detecting more complex, dynamic actions remains limited. Future research should focus on compiling more diverse datasets encompassing various complex and dynamic physical activities. The investigation of advanced DL algorithms incorporating attention mechanisms and memory networks could also improve the model's capacity to capture and interpret fine-grained movement details.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author Shamsulariffin Bin Samsudin on reasonable request via e-mailshamariffin@upm.edu.my.

Received: 3 November 2024; Accepted: 9 May 2025
Published online: 16 May 2025

References

1. Jaouedi, N., Boujnah, N. & Bouhlel, M. S. A new hybrid deep learning model for human action recognition. *J. King Saud Univ. Comput. Inform. Sci.* **32** (4), 447–453 (2020).
2. Gupta, N. et al. Human activity recognition in artificial intelligence framework: a narrative review. *Artif. Intell. Rev.* **55** (6), 4755–4808 (2022).
3. Pareek, P. & Thakkar, A. A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. *Artif. Intell. Rev.* **54** (3), 2259–2322 (2021).
4. Dong, Z. & Sha, N. Research on the Current Situation and Countermeasures of Cultivating Talents in Recreational Sports Under the Perspective of Artificial Intelligence. *Appl. Math. Nonlinear Sci.* (2024).
5. Guangde, Z. Physical education and emergency response system using deep learning: A step toward sustainable development of physical education environment. *Front. Environ. Sci.* **10**, 974291 (2022).
6. Li, F. Creation of deep learning scenarios in the network teaching of physical education technical courses. *Scalable Comput. Pract. Exp.*. **25** (1), 271–284 (2024).
7. Gu, F. et al. A survey on deep learning for human activity recognition. *ACM Comput. Surv. (CSUR)*. **54** (8), 1–34 (2021).
8. Kong, Y. & Fu, Y. Human action recognition and prediction: A survey. *Int. J. Comput. Vis*. **130** (5), 1366–1401 (2022).
9. Wan, S. et al. Deep learning models for real-time human activity recognition with smartphones. *Mob. Netw. Appl.* **25** (2), 743–755 (2020).
10. Sun, Z. et al. Human action recognition from various data modalities: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **45** (3), 3200–3225 (2022).
11. Khan, M. A. et al. A resource conscious human action recognition framework using 26-layered deep convolutional neural network. *Multimed. Tools Appl.* **80**, 35827–35849 (2021).
12. Wei, S. et al. Exploring the application of artificial intelligence in sports training: a case study approach. *Complexity* **2021** (1), 4658937 (2021).
13. Huang, A. A. & Huang, S. Y. Use of machine learning to identify risk factors for insomnia. *PloS One*. **18** (4), e0282622 (2023).
14. Huang, A. A. & Huang, S. Y. Quantification of the relationship of pyridoxine and spirometry measurements in the united States population. *Curr. Dev. Nutr.* **7** (8), 100078 (2023).
15. Huang, A. A. & Huang, S. Y. Shapely additive values can effectively visualize pertinent covariates in machine learning when predicting hypertension. *J. Clin. Hypertens.* **25** (12), 1135–1144 (2023).
16. Liao, S. & Fu, C. The optimization of youth football training using deep learning and artificial intelligence. *Sci. Rep.* **15** (1), 8190 (2025).
17. Huang, A. A. & Huang, S. Y. Increasing transparency in machine learning through bootstrap simulation and shapely additive explanations. *PLoS One*. **18** (2), e0281922 (2023).
18. Demrozi, F. et al. Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access.* **8**, 210816–210836 (2020).
19. Agarwal, P. & Alam, M. A lightweight deep learning model for human activity recognition on edge devices. *Procedia Comput. Sci.* **167**, 2364–2373 (2020).
20. Ferrari, A. et al. On the personalization of classification models for human activity recognition. *IEEE Access.* **8**, 32066–32079 (2020).
21. Dua, N., Singh, S. N. & Semwal, V. B. Multi-input CNN-GRU based human activity recognition using wearable sensors. *Computing* **103** (7), 1461–1478 (2021).
22. Mazzia, V. et al. Action transformer: A self-attention model for short-time pose-based human action recognition. *Pattern Recogn.* **124**, 108487 (2022).
23. Mekruksavanich, S. & Jitpattanakul, A. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. *Sensors* **21** (5), 1636 (2021).
24. Yadav, S. K. et al. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowl. Based Syst.* **223**, 106970 (2021).

25. Zhan, C. & Cui, P. Predicament and strategy of campus football teaching under the background of artificial intelligence and deep learning. *J. Comput. Methods Sci. Eng.* **23** (5), 2437–2449 (2023).
26. Long, Y. Research on Football Sports Classification System Based on Artificial Intelligence. *Proceedings of the 2023 International Conference on Information Education and Artificial Intelligence*. 703–707. (2023).
27. Beddiar, D. R. et al. Vision-based human activity recognition: a survey. *Multimed. Tools Appl.* **79** (41), 30509–30555 (2020).
28. Song, Y. F. et al. Richly activated graph convolutional network for robust skeleton-based action recognition. *IEEE Trans. Circuits Syst. Video Technol.* **31** (5), 1915–1925 (2020).
29. Sherafat, B. et al. Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. *J. Constr. Eng. Manag.* **146** (6), 03120002 (2020).
30. Patel, C. I. et al. Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences. *Sensors* **20** (24), 7299 (2020).
31. Garcia-Gonzalez, D. et al. A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors* **20** (8), 2200 (2020).
32. Guha, R. et al. CGA: A new feature selection model for visual human action recognition. *Neural Comput. Appl.* **33**, 5267–5286 (2021).
33. Gupta, R. et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers*. **25**, 1315–1360 (2021).
34. Song, Y. F. et al. Constructing stronger and faster baselines for skeleton-based action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **45** (2), 1474–1488 (2022).
35. Shu, X. et al. Expansion-squeeze-excitation fusion network for elderly activity recognition. *IEEE Trans. Circuits Syst. Video Technol.* **32** (8), 5281–5292 (2022).
36. Ullah, I. & Mahmoud, Q. H. A two-level flow-based anomalous activity detection system for IoT networks. *Electronics* **9** (3), 530 (2020).
37. Jegham, I. et al. Vision-based human action recognition: an overview and real world challenges. *Forensic Sci. Int. Digit. Invest.* **32**, 200901 (2020).
38. Xu, Y. et al. Machine learning in construction: from shallow to deep learning. *Dev. Built Environ.* **6**, 100045 (2021).
39. Nahavandi, D. et al. Application of Artificial Intelligence in Wearable Devices: Opportunities and Challenges. *Comput. Methods Progr. Biomed.*, 213106541 (2022).
40. Yang, Y., Zhuang, Y. & Pan, Y. Multiple knowledge representation for big data artificial intelligence: framework, applications, and case studies. *Front. Inform. Technol. Electron. Eng.* **22** (12), 1551–1558 (2021).
41. Pallathadka, H. et al. Impact of machine learning on management, healthcare and agriculture. *Mater. Today Proc.* **80**, 2803–2806. (2023).
42. Huynh-The, T. et al. Artificial intelligence for the metaverse: A survey. *Eng. Appl. Artif. Intell.* **117**, 105581 (2023).
43. Rao, H. et al. Augmented skeleton based contrastive action learning with momentum Lstm for unsupervised action recognition. *Inf. Sci.* **569**, 90–109 (2021).
44. Segalin, C. et al. The mouse action recognition system (MARS) software pipeline for automated analysis of social behaviors in mice. *Elife* **10**, e63720 (2021).
45. Shi, L. et al. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks. *IEEE Trans. Image Process.* **29**, 9532–9545 (2020).
46. Zhang, J. & Tao, D. Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things. *IEEE Internet Things J.* **8** (10), 7789–7817 (2020).
47. Shu, X. et al. Host–parasite: graph LSTM-in-LSTM for group activity recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **32** (2), 663–674 (2020).

Author contributions

Xueyan Ji: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation Shamsulariffin Bin Samsudin: writing—review and editing, visualization, supervision, project administration, funding acquisitionMuhammad Zarif Bin Hassan: methodology, software, validation Noor Hamzani Farizan: formal analysis, investigation, resourcesYubin Yuan: visualization, supervisionWang Chen: visualization, supervision, project administration.

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

The studies involving human participants were reviewed and approved by Department of Sports Studies, Facility of Educational Studies, Universiti Putra Malaysia Ethics Committee (Approval Number: 2022.200219819996). The participants provided their written informed consent to participate in this study. All methods were performed in accordance with relevant guidelines and regulations.

Additional information

Correspondence and requests for materials should be addressed to S.B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.