SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(3), 484.
https://doi.org/10.62617/mcb484

Article

Sports training injury risk assessment combined with dynamic analysis algorithm

Zhihong Hou, Yuan Xue*

Physical Education Department, Qinhuangdao vocational and technical college, Qinhuangdao 066100, Hebei, China
*** Corresponding author:** Yuan Xue, 13803358622@163.com

CITATION

Hou Z, Xue Y. Sports training injury risk assessment combined with dynamic analysis algorithm. Molecular & Cellular Biomechanics. 2024; 21(3): 484. https://doi.org/10.62617/mcb484

ARTICLE INFO

Received: 9 October 2024
Accepted: 23 October 2024
Available online: 18 November 2024

COPYRIGHT

Copyright © 2024 by author(s). *Molecular & Cellular Biomechanics* is published by Sin-Chn Scientific Press Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/by/4.0/

Abstract: To explore the application of dynamic analysis algorithm in sports training injury risk assessment, this paper takes the Spatio-Temporal Graph Convolutional Network (ST-GCN) as the main algorithm, and introduces the Adaptive Graph Convolution Module (AGCM) and Residual Channel Attention Module (RCAM). ST-GCN is improved to form AGCM + RCAM-ST-GCN (ARST-GCN) motion posture recognition algorithm. Meanwhile, combined with the extreme gradient boosting (XG Boost), the final physical training injury risk assessment model is formed. The performance of the improved ARST-GCN and the proposed damage risk assessment model is verified by experiments. The results show that ARST-GCN, which combines AGCM and RCAM modules, performs best in all indicators. Compared with ST-GCN, the accuracy rate is increased by 1.94% and the $F1$ value is increased by 4.3%. In addition, in the performance comparison of different sports injury risk models, the recall rate and $F2$ value of XGBoost are 0.937 and 0.893, respectively, and the overall performance is the best, indicating that XGBoost has significant advantages in dealing with sports injury risk assessment (SIRA) tasks. The research results provide theoretical basis and practical reference for injury prevention in sports training, and help to improve the accuracy and reliability of SIRA.

Keywords: physical training; injury risk; spatio-temporal graph convolutional network; adaptive graph convolution module; residual channel attention module

1. Introduction

With the increasing importance of sports in the field of health and competition, the scientific and meticulous management of sports training has gradually become the key to improving sports performance and ensuring athletes' health [1,2]. Accurate evaluation and prevention of sports injuries has become one of the focuses of sports science research. The major components of traditional sports injury risk assessment (SIRA) systems are athlete self-reports and expert expertise. Although these methods can reveal the potential risks of sports injuries to a certain extent, they often lack real-time and accuracy [3]. Furthermore, these approaches have limited practical relevance because they are unable to anticipate and alert to the possibility of injury early in the training phase [4].

SIRA has made extensive use of big data and machine learning techniques in recent years due to the quick advancements in sensing and data processing technologies [5]. The foundation of assessing the risk of sports injuries is motion posture identification. As an advanced dynamic analysis algorithm, Spatio-temporal graph convolutional network (ST-GCN) shows its unique advantages in processing graph data containing spatio-temporal information [6]. ST-GCN can effectively

capture the dynamic changes in the process of sports by constructing spatio-temporal graph structure and modeling the spatial and temporal characteristics of sports data [7].

Adaptive graph convolutional module (AGCM) and residual channel attention module (RCAM) are introduced in this research with the goal of optimizing the classic ST-GCN, which is then used to the SIRA model. The paper offers a novel technical approach to injury risk assessment in sports training, as well as methodological guidance and theoretical backing for related domains of study and practice.

The main contributions of this paper are as follows:

Firstly, an improved ST-GCN is proposed. By introducing AGCM and RCAM, the accuracy and robustness of motion posture recognition are improved.

Secondly, the improved ST-GCN is applied to the SIRA model, which provides a more accurate SIR prediction method.

Finally, the effectiveness of the proposed model in sports posture recognition and injury risk assessment is verified by experiments, which provides technical support and theoretical basis for scientific management of sports training and sports injury prevention.

2. Related work

SIRA aims to predict and prevent potential sports injuries by analyzing athletes' sports data. At present, research has made progress in sports training injury risk assessment. An innovative dual-feature fusion neural network model is proposed by Meng and Qiao. The issue of feature loss was resolved by applying 1 × 1 convolution and hyperlink to create a dual fusion structure, which was then used to assess sports injuries [8]. Their study provided a new idea for sports injury prediction and highlighted the importance of feature fusion in deep learning. In their discussion of the use of machine learning in football injury risk assessment, Nassis et al. noted that while machine learning's capacity for injury prediction was currently limited, it could aid in determining the early risk of musculoskeletal injuries [9]. This showed that although technology was progressing, the accuracy and reliability of machine learning model needed to be further improved. A recurrent neural network (RNN) model was created by Dhanke et al. to examine the impact of sports training on injuries sustained by players. By gathering and examining training data from athletes, the model assessed and forecasted sports injuries [10]. This study emphasized the importance of dynamic training data for injury risk assessment and laid a foundation for future research. Rebelo et al. evaluated the application of technology in sports training and injury prevention, and found that wearable devices and power boards were widely used in monitoring athletes' performance and injury prediction. Especially, the sports load data was very important for customizing training load [11]. This discovery emphasized the practical application of technology in sports science and provided theoretical support for personalized training. Li and Zhu put forward a sports injury risk analysis system based on neural network, which combined blockchain and Internet of Things to solve the common injury problems in athletes' training. Through multi-sensor data fusion technology, the system could quickly identify the damage location within 0.2 seconds. The recovery rate reached 94.39%, which significantly improved the accuracy and stability of damage monitoring [12]. Their study showed how advanced

technology could be applied in sports medicine to improve the response speed and accuracy.

In addition, the existing motion posture recognition algorithms have also made remarkable progress. To detect the movement posture, Hu et al. introduced a unique multi-scale time sampling module and depth spatio-temporal feature extraction module, which increased the spatio-temporal feature extraction and the feature map's receptive field [13]. The innovation of this method lied in its multi-level consideration of feature extraction, which provided a new methodology for complex motion recognition. Wang used a deep learning technique based on convolutional neural networks (CNNs) to address the issue of traditional feature extraction methods' reliance on manual design. This allowed for the identification of key posture time through cosine similarity in fitness scenes using a multi-scene action similarity analysis algorithm based on human joint points [14]. Their study provided a new way for motion posture analysis based on visual data, and highlighted the value of action similarity in the recognition process. Arab et al. discussed the feasibility and effectiveness of using 24 GHz Doppler radar to detect and extract human motion signals through theoretical analysis, simulation and experimental verification, and applied dual-channel CNN to learn high-level features, achieving 98.85% accuracy of motion classification [15]. The findings showed that the potential of radar technology in motion analysis opened a new direction for the application of sensor technology in the future. Dong and Wang proposed a method for recognizing athletes' gestures based on improved deep neural network, designed a standard for recognizing sports gestures, and extracted fusion features through image processing. The experimental results showed that this method had a high recognition rate and the recognition time was only 1.2 seconds [16]. This achievement emphasized the importance of rapid reaction time in athletes' training monitoring. Li and Boers studied a method of human motion recognition in dance video based on pose estimation algorithm. The experimental results showed that the average recognition rate of this algorithm (77.95%) was slightly higher than that of other algorithms (such as deep learning algorithm 76.23%), which showed its application potential in dance teaching and motion analysis [17]. Their study provided technical support for traditional dance teaching methods and showed the practicability of computer vision in the art field.

Sports posture recognition algorithms and sports training injury risk assessment have made some progress thus far. The motion posture recognition algorithm greatly enhances the ability to recognize complicated motion postures, and the SIRA approach improves the assessment's accuracy and real-time performance. However, these methods still have shortcomings in dealing with diverse sports scenes and individual differences, and need to be further improved and optimized. In this paper, a SIRA model combined with an improved ST-GCN is proposed. By introducing AGCM and RCAM, the applicability and robustness of the model in different sports scenes are improved, which provides a new solution and theoretical support for sports training injury risk assessment.

3. SIRA model combined with ST-GCN

The dynamic analysis algorithm is to model and analyze the time series data, and extract the dynamic characteristics from it to reveal the law of the object changing with time [18]. Sports training injury risk assessment is a specific application for this type of algorithm since sports activities contain intricate spatial and temporal dynamic features. ST-GCN, as a typical dynamic analysis algorithm, can effectively capture and deal with the changes in spatial and temporal dimensions during the movement process, and provide a deep understanding of the movement pattern [19]. The paper proposes a novel SIRA model to assist the safety of sports training by integrating AGCM and RCAM to improve ST-GCN.

3.1. ST-GCN

ST-GCN is a deep learning model that combines Graph Convolutional Network (GCN) and Temporal Convolution Network (TCN). By constructing a graph structure containing the relationship between space and time, the spatio-temporal characteristics are extracted by graph convolution operation, and the dynamic process is modeled and analyzed. The basic principle of ST-GCN is based on transforming time series data into graph structure to model spatial and temporal characteristics [20]. The initial step in the data construction process for human motion detection in ST-GCN is to create a spatio-temporal graph. The nodes of the network represent the important locations on the human skeleton, while the edges show the connections between these important locations. Assuming that there are T frames of motion data, each frame contains n skeleton joint points, the graph $G = (P, E)$ can be expressed as a graph, where P is a node set and E is an edge set. Specifically, the node set P includes joint points in all frames, and is defined as:

$$P = \{p_{t,i} | t = 1,2,\cdots,T; i = 1,2,\cdots,n\} \tag{1}$$

Edge set E consists of two parts: spatial edge E_s and time edge E_t. Space edge E_s describes the connection between nodes in the same frame, while time edge E_t describes the connection between the same joint points across frames. The equation is expressed as:

$$E = E_s \cup E_t \tag{2}$$

The spatial edge E_s is composed of the spatial connection set S in the graph:

$$E_s = \{(p_{t,i}, p_{t,j}) | (i,j) \in S\} \tag{3}$$

Time edge E_t is the edge connecting the same joint points in adjacent frames:

$$E_t = \{(p_{t,i}, p_{t+1,i}) | t = 1,2,\cdots,T-1\} \tag{4}$$

In ST-GCN, spatio graph convolution and temporal convolution are key operations. By merging data from each node and its surrounding nodes, the spatial graph convolution operation convolves the spatial graph in each frame with the aim of updating each node's properties [21]. To gain a better understanding of the convolution product of spatio graph, the conventional two-dimensional convolution process is compared. Its form for the conventional two-dimensional convolution process is:

$$f_{out}(x) = \sum_{k=1}^{H} \sum_{w=1}^{H} \omega(k, w) \times f_{in}(x - k, y - w) \quad (5)$$

f_{in} is the input characteristic graph. f_{out} is the output characteristic graph. x is the abscissa of a specific position of the input feature map, and y is the ordinate of a specific position of the input feature map. The size of convolution kernel is $H \times H$. $\omega(k, w)$ is the weight value of convolution kernel, and the weight at the height of k and the width of w.

For the convolution operation in graph data, the convolution process becomes more complicated because of the irregular neighborhood structure of nodes. The neighborhood set V of nodes is usually expressed as:

$$V(p_{t,i}) = \{(p_{t,j} | d(p_{t,i}, p_{t,j}) \leq D\} \quad (6)$$

$d(p_{t,i}, p_{t,j})$ represents the distance between nodes $p_{t,i}$ and $p_{t,j}$. D is the distance threshold of spatial neighborhood.

In ST-GCN, the modeling of time dimension is completed by time convolution network, which can capture the dynamic changes in time series data [22]. The spatio-temporal graph convolution operation extends the spatial neighborhood of each node to the time dimension, which is defined as:

$$V_{ST}(p_{t,i}) = \{(p_{t,j}, p_{a,b}) | d(p_{t,i}, p_{t,j}) \leq D \; and \; |t - a| \leq R\} \quad (7)$$

$V_{ST}(p_{t,i})$ is a spatio-temporal neighborhood, that is, a set of nodes adjacent to node $p_{t,i}$ in time and space. $p_{a,b}$ is the b-th node in the a-th frame, and it is a node in a different time frame from node $p_{t,i}$. $d(p_{t,i}, p_{t,j})$ is the spatial distance, and $|t - a|$ is the time span. R is the scale of the time convolution kernel, which determines the scope of the time neighborhood, that is, the distance $|t - a|$ between two nodes in the time dimension must be less than or equal to R before they are adjacent. The characteristics of node $p_{t,i}$ in the spatio-temporal neighborhood can be updated by combining the characteristics of nodes in the spatio-temporal neighborhood, such as:

$$f_{ST}(p_{t,i}) = \sum_{(p_{t,j}, p_{a,b}) \in V_{ST}(p_{t,i})} \omega(p_{t,i}, p_{t,j}, p_{a,b}) \times f(p_{t,j}, p_{a,b}) \quad (8)$$

$f_{ST}(p_{t,i})$ is the updated feature of node $p_{t,i}$ in the spatio-temporal neighborhood. ω is the weight function of spatio-temporal convolution kernel, and $f(p_{t,j}, p_{a,b})$ is the nodes $p_{t,j}$ and $p_{a,b}$.

3.2. Improved ST-GCN

To improve the accuracy and robustness of ST-GCN in motion recognition, AGCM and RCAM are introduced to improve the original ST-GCN, and the improved ST-GCN is called ARST-GCN (AGCM + RCAM − ST-GCN). While RCAM weights the feature channels, AGCM can dynamically modify the adjacency matrix, improving the model's capacity to extract important data. **Figure 1** depicts the ARST-GCN model's architecture.

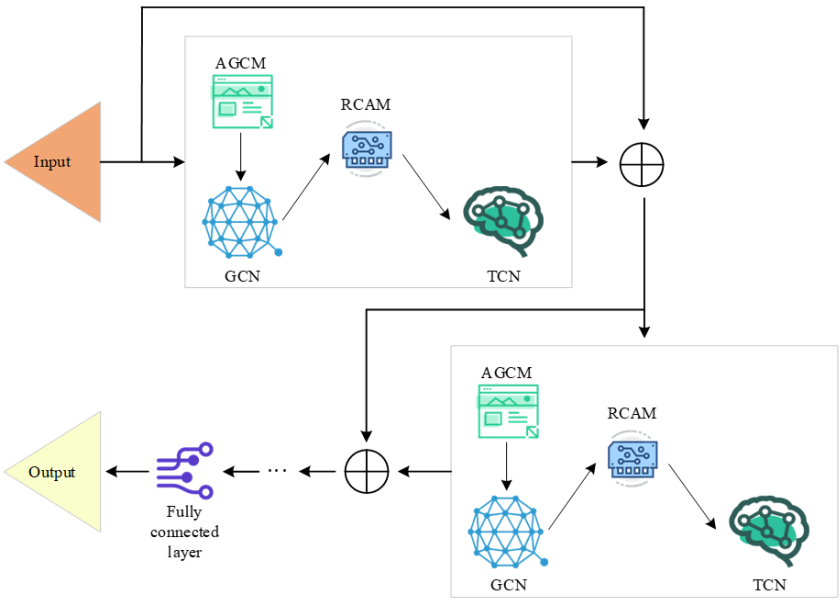

Figure 1. ARST-GCN model architecture diagram.

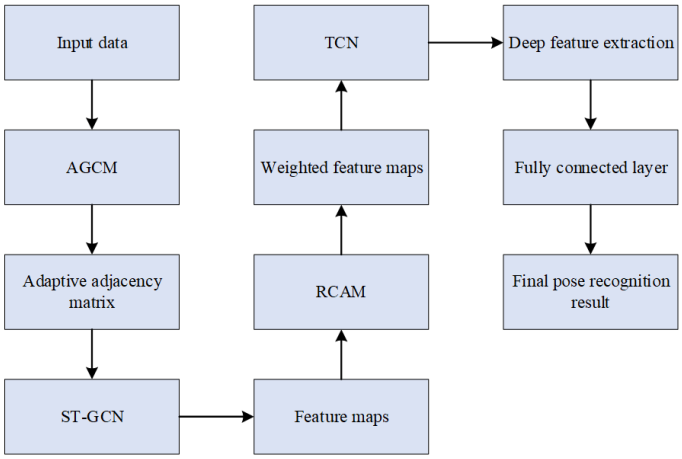

Figure 2. Data processing flow in ARST-GCN model.

In ARST-GCN model, the input data is preprocessed by AGCM module. AGCM adjusts the adjacency matrix according to the dynamic characteristics of input features to optimize the graph structure. In this process, AGCM integrates data-driven information feedback into its calculation, and enhances the expressive ability of GCN by updating adjacency matrix. The processed feature map is then transferred to ST-GCN. In this process, the adaptive adjacency matrix generated by AGCM can significantly improve ST-GCN's ability to capture motion patterns. On this basis, RCAM module is introduced to further optimize the model performance. RCAM is closely combined with ST-GCN, which receives the feature map from ST-GCN, calculates the channel weights, and adjusts the features. RCAM processes features by embedding global information to ensure that key motion features are strengthened. This weighting operation not only improves the sensitivity to subtle movement changes, but also enhances the model's ability to capture key information. Finally, the feature map optimized by AGCM and RCAM modules is extracted by TCN to ensure the complete retention of time dimension information. In the output stage, all the

processed feature maps are fused, and the final motion posture recognition result is generated through the fully connected layer. The whole process can be summarized as the flow chart shown in **Figure 2**.

(1) Adaptive graph convolution module

The fundamental principle of AGCM is to use data-driven mode to adaptively modify the graph structure to optimize it based on the properties of real data, hence increasing the model's flexibility and accuracy [23]. In the basic ST-GCN, let A be the adjacency matrix representing the physiological connection between the key points of human bones, and the basic graph convolution operation can be expressed as:

$$f_{out} = \sum_{k=1}^{K} W_k f_{in} A_{norm}^k \tag{9}$$

W_k is the trainable weight matrix. A_{norm}^k is the normalized adjacency matrix. f_{in} and f_{out} are the input and output characteristic graphs respectively. To introduce adaptability, this paper proposes a data-driven adaptive graph convolution product, which is expressed as:

$$f_{out} = \sum_{k=1}^{K} W_k f_{in} (Z_k + C_k) \tag{10}$$

Z_k is a data-guided graph adjacency matrix, and its initial value is the parameters of the basic adjacency matrix A_k. In the whole network training process, the parameters of Z_k are updated, so that it can adjust the graph structure adaptively under the guidance of training data. Compared with the fixed basic adjacency matrix A_k, the adaptive graph adjacency matrix Z_k can automatically adjust the connection relationship and strength between nodes according to the data, and even generate connections that do not exist in the original physiological structure, thus describing the dynamic characteristics of the human skeleton more flexibly [24,25]. Specifically, the updating process of Z_k is as follows:

$$Z_k = A_k + Q_k \tag{11}$$

Q_k is a learnable mask matrix. By learning the connection strength between different joints during training, the original adjacency matrix is optimized.

In addition, C_k is also an adjacency matrix of adaptive graph, which captures the interaction between two nodes by using normalized Gaussian embedding function by referring to non-local attention mechanism, and is defined as follows:

$$C_{ij} = \frac{e^{\varphi(p_i)^T \times \theta(p_j)}}{\sum_{j=1}^{N} e^{\varphi(p_i)^T \times \theta(p_j)}} \tag{12}$$

$\varphi(*)$ and $\theta(*)$ are mapping functions. p_i and p_j are two nodes in the same frame. By calculating their similarity, the connection relationship and strength between nodes are determined. The specific calculation steps of this process are as follows: first, the input feature map f_{in} is mapped to a new feature space, and the feature map with the size of C × T × N is obtained. Then, two embedding functions φ and θ are used to map it, and data with sizes of N × C × T and C × T × N are obtained. Next, the two characteristic matrices are multiplied to obtain a similarity matrix C_k with the size of

N × N. C_{ij} represents the similarity between nodes p_i and p_j. Because normalized Gaussian embedding is equivalent to softmax operation, the calculation method of matrix C_k can also be expressed as:

$$C_k = \text{softmax}(f_{in}^T W_{\theta k}^T W_{\varphi k} f_{in}) \tag{13}$$

(2) Residual channel attention module

RCAM is introduced into ST-GCN, and the network's ability to capture motion features is improved through adaptive correction and enhancement of different channel features [26]. This module's primary goal is to give the feature information of various channels corresponding weights to emphasize key aspects and increase the model's accuracy and resilience [27]. **Figure 3** depicts the RCAM's construction.

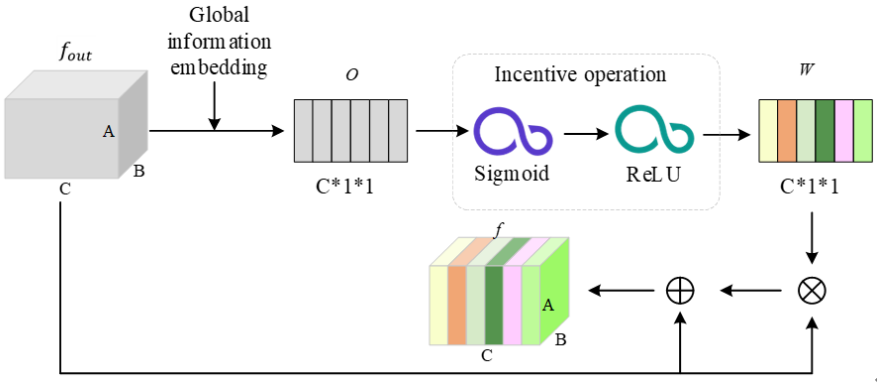

Figure 3. Residual channel attention module.

Firstly, the initial spatial feature map f_{out} is embedded with global information, which is realized by global average pooling. Specifically, the feature map with the size of C × A × B is compressed into a vector O with the size of C × 1 × 1, and the equation is as follows:

$$O_c = F_{sq}(L_c) = \frac{1}{A \times B} \sum_{i=1}^{A} \sum_{j=1}^{B} L_c(i,j) \tag{14}$$

F_{sq} is a global information embedding function, and the feature map is compressed from three dimensions to one dimension by global average pooling. Its function is to compress the two-dimensional feature map of each channel into a scalar, which represents the global characteristics of the channel. $L_c(i,j)$ represents the characteristic value of the output characteristic graph at the position (i,j) on channel c.

Next, the vector O is excited by two fully connected layers. Firstly, the vector O is mapped to a $\frac{C}{s}$ dimension vector through the first fully connected layer, and ReLU activation is performed to get the intermediate vector O_1. Then, O_1 is mapped back to the C-dimensional vector through the second fully connected layer, and Sigmoid activation is carried out to obtain the weight vector W, and the specific calculation process is as follows:

$$W = F_{ex}(O) = \sigma(Q_2(\delta(Q_1 O))) \tag{15}$$

F_{ex} is an excitation operation function, and the compressed vector is transformed nonlinearly through two fully connected layers. Its function is to map the compressed vector O into a weight vector W with the same number of channels as the input feature map through the combination of two fully connected layers and activation functions to adjust the importance of each channel. $Q_1 \in \mathbb{R}^{\frac{C}{s} \times C}$ and $Q_2 \in \mathbb{R}^{C \times \frac{C}{s}}$ are the parameter matrices of two fully connected layers respectively. σ represents Sigmoid activation function and δ represents ReLU activation function.

The weight vector W is multiplied by the initial spatial feature map f_{out} element by element, and then the residual connection is made with the feature map before correction, and the final output feature map f is obtained by addition. The equation is as follows:

$$f = f_{out} \odot W + f_{out} \tag{16}$$

To sum up, AGCM dynamically adjusts the adjacency matrix in a data-driven way, so that the network can flexibly adapt to the temporal and spatial characteristics of the data. This adaptive adjustment not only optimizes the representation ability of graph structure, but also enhances the robustness of the model to dynamic data. On the other hand, RCAM ensures the prominent display of important features by weighting the features of each channel, thus improving the sensitivity of the model to subtle movement changes. Combining the advantages of these two modules, the improved ST-GCN shows higher accuracy and robustness when dealing with complex motion data, and can capture the subtle changes in human motion more effectively.

3.3. Risk assessment model of sports injury based on ARST-GCN

The enhanced ARST-GCN motion posture recognition algorithm is used to the model to increase the SIRA's accuracy and resilience. **Figure 4** illustrates the structure of the sports risk assessment model.

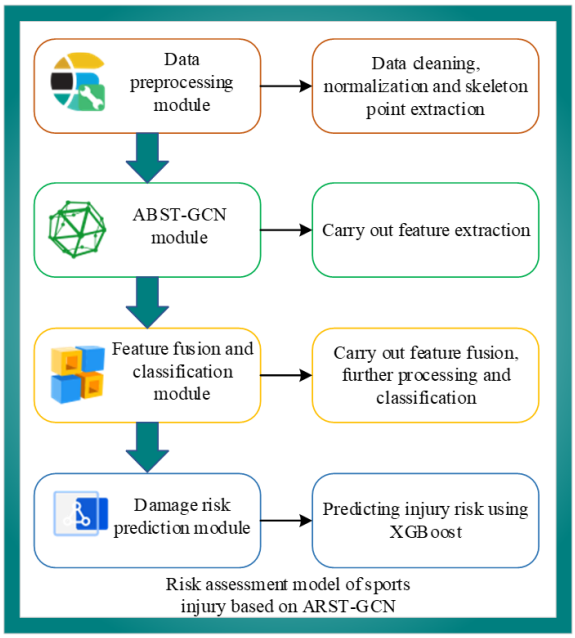

Figure 4. SIRA model based on ARST-GCN.

This model mainly includes the following main modules: data preprocessing module, improved ST-GCN module, feature fusion and classification module, and damage risk prediction module. Firstly, the motion data is preprocessed, including data cleaning, normalization and skeleton point extraction. This process ensures the consistency and quality of input data, and lays a foundation for subsequent feature extraction and modeling. Secondly, ARST-GCN is used to extract features. The spatio-temporal features extracted by ARST-GCN are fused, and a multi-layer neural network is constructed for further processing and classification of features. Through the full connected layer and activation function, the module performs dimensionality reduction and nonlinear transformation on the extracted features, and finally outputs the classification results of motion postures. Finally, the extreme gradient boosting (XG Boost) technique is used to predict the injury risk of athletes during the training process by combining the associated factors of sports injuries with the movement patterns that are identified, utilizing the output results of the feature fusion and classification module [28]. The SIRA model based on ARST-GCN can efficiently identify and assess potential injury risks during sports activities and offer real-time risk warning and assessment recommendations thanks to the collaborative efforts of the above modules.

4. Model performance verification

4.1. Experimental design

This experiment is divided into two parts.

Experiment 1: Performance verification of ARST-GCN algorithm. This experiment is conducted on Nanyang Technical University Red Green Blue + Depth (NTU RGB + D) dataset, which is a standard data set widely used in motion posture recognition. The NTU RGB + D dataset contains 60 action categories with a total of 56,880 action samples, and records the RGB video, depth image and skeleton data of each action. Each sample in the dataset consists of multiple skeleton points, and the coordinates of these skeleton points in three-dimensional space are used to model the posture and movements of athletes. To ensure the fairness and reliability of model evaluation, the dataset is divided into training set, verification set and test set according to the ratio of 8:1:1.

In the process of hyperparameter optimization, firstly, several key parameters are optimized by grid search method. These parameters include learning rate, weight attenuation coefficient and batch size. Selections are made from a predetermined range. For example, the learning rate is chosen from 0.01 to 0.1, the weight decay coefficient is chosen from 1e-4 to 1e-2, and the batch size is chosen from 16 to 64. The optimal combination of hyperparameter is determined by evaluating the performance of the model on the validation set. In addition, an early stopping strategy is used to avoid the overfitting phenomenon. During the training process, the training will be stopped when the loss on the validation set does not improve in 10 consecutive epochs. This strategy effectively balances the training time and performance of the model.

The experimental running environment and parameter settings are shown in **Table 1**.

Table 1. Experimental environment and parameter settings.

Hardware/parameter name	Parameter/value
Operating system	Windows10
CPU	AMD R7-5800H
Display card	NVIDIA GeForce RTX3090
Hard disc	512G SSD
Deep learning framework	Pytorch 1.7.0
Initial learning rate	0.1
Epoch	50
Data volume of each batch	32
Weight attenuation coefficient	1 × e−4

The experiment's evaluation indices for motion posture recognition are the accuracy rate, precision rate, recall rate, and *F*1 value. Let *TN* be the number of samples that are accurately classified as belonging to this category. *FP* stands for the number of samples that other categories incorrectly anticipated to fall into this category. The number of samples that do not belong in this category and are expected to not belong in this category is represented by *TP*, while the number of samples that this category label predicts to belong in other categories is represented by *FN*. The accuracy calculation equation is:

$$Acc = \frac{TN + TP}{TN + FP + TP + FN} \tag{17}$$

*F*1 value is the harmonic average of precision and recall. Precision represents the proportion of the sample of the predicted category that belongs to the category, and its calculation expression is shown in Equation (18).

$$Pre = \frac{TN}{TN + FP} \tag{18}$$

Recall indicates the correct sample ratio among all samples belonging to this category, which is calculated by Equation (19).

$$Rec = \frac{TN}{TN + FN} \tag{19}$$

The calculation expression of the *F*1 value is shown in the public notice (20).

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec} \tag{20}$$

Experiment 2: Performance verification of SIRA model based on ARST-GCN. 20 male volunteers are selected to carry out SIRA experiments. All participants need to sign an informed consent form before the start of the experiment, clearly informing them of the purpose, process and potential risks of participating in the study. At the same time, participants have the right to withdraw from the experiment at any time without any consequences. All the collected data are anonymously processed in the process of storage and analysis to ensure that the personal identity information of the participants is not leaked, including the de-identification of data such as sports training load, health status and injury records. In addition, all data are stored on a secure server,

and access is limited to the research team members. Through the above measures, the personal privacy and data security of participants are fully respected and protected. The average age of these participants is 20 years old, and the data of exercise training load, subjective perception of health, exercise quality evaluation and injury record of these 20 volunteers are monitored during the six-month period. Data collection is displayed in **Table 2**.

Table 2. Data collection of experiment 2.

Data classification	Quantization method	Quantitative index	Explain
Sports training load data	Quantitative index of training load	Subjective physical sensation scale	Assess participants' subjective feelings about physical load.
	Cumulative training load index	Short-term cumulative load	Record the training load within 7 days.
		Long-term cumulative load	Record the training load within 28 days.
	Trend index of training load change	Monotonicity of training load	Evaluate the changing law of training load
		Short-term long-term load ratio	Ratio of short-term load to long-term load
Subjectively perceived health data	Health status survey scale		The health status of participants is obtained through questionnaire survey.
Sports quality evaluation data	Test results of squats, push-ups, 100 m sprints, etc.		Record the test results of these trainings.
Damage record data	Medical examination	Record the damage location	Record the location of the damage.
		Damage property	Record the nature of the injury (such as tear, contusion, etc.)
		Damage type	Record the types of injuries (such as fractures, sprains, etc.)
		Damage occurrence	Record the occurrence of injuries (e.g. during training, competition, etc.)

The dataset is split at random into a training set and a verification set with a 9:1 ratio for the experiment. The model's precision, recall rate, and *F*2 value are utilized as performance metrics. Recall rates must be increased while maintaining precision because the cost of misdiagnosing an injury is far higher than that of correctly diagnosing it. The following calculation approach yields an *F*2 number that accurately represents the prediction model's overall performance on both the majority and minority samples:

$$F2 = \frac{5 \times Pre \times Rec}{4 \times Pre + Rec} \tag{21}$$

4.2. Performance verification of ARST-GCN motion recognition algorithm

(1) Ablation experiment

To test the performance of ARST-GCN model, several groups of ablation experiments are designed to verify the effectiveness of AGCM and RCAM modules.

Firstly, the performance of AGCM module is verified, in which there are two graph adjacency matrices Z_k and C_k, which are verified respectively. The

experimental results of ablation based on the adjacency matrix Z_k of adaptive graph are shown in **Figure 5**.

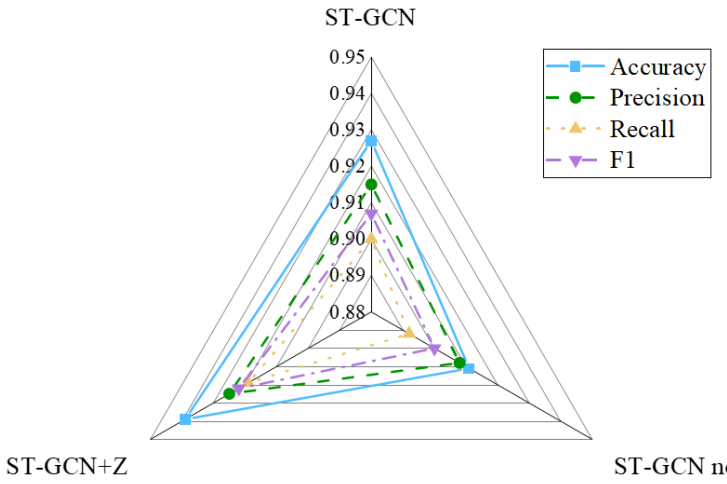

Figure 5. Experimental results of ablation based on graph adjacency matrix z k .

In **Figure 5**, ST-GCN noQ represents the case where the mask matrix Q_k is removed in the ST-GCN model, and ST-GCN + Z represents the case where only the graph adjacency matrix Z_k is used. **Figure 5** shows that the performance of the model without the mask matrix Q_k is the worst, indicating the importance of the mask matrix. The model using graph adjacency matrix Z_k performs better, and its accuracy, precision, recall and $F1$ value are 0.939, 0.925, 0.919 and 0.922, respectively. Compared with the basic ST-GCN model, the model using Z_k has achieved remarkable improvement in all evaluation indexes, and the accuracy, precision, recall and $F1$ value have increased by 1.29%, 1.09%, 2.11% and 1.65% respectively. This fully demonstrates the effectiveness of graph adjacency matrix Z_k of AGCM module in feature capture and information flow. This is because through the graph adjacency matrix Z_k, the model can dynamically adjust the connection relationship and strength between nodes according to the characteristics of training data, thus optimizing the movement posture recognition process. Z_k not only improves the accuracy of the model, but also enhances its ability to identify key information, ensuring that different categories can be effectively distinguished in complex motion postures.

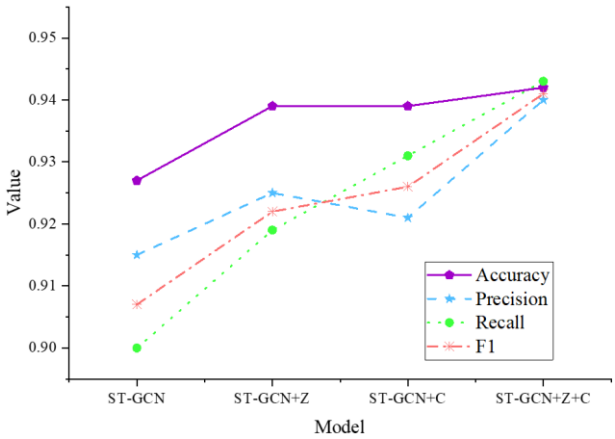

Figure 6. Experimental results of ablation based on graph adjacency matrix *c k*.

The experimental results of ablation based on graph adjacency matrix C_k are shown in **Figure 6**.

In **Figure 6**, the ST-GCN + Z + C model combined with the adjacency matrix of two graphs has the best performance in accuracy, precision, recall and $F1$ value, with accuracy of 0.942, precision of 0.94, recall of 0.943 and $F1$ value of 0.941. Compared with the model with only one graph adjacency matrix, the accuracy, precision and $F1$ value are improved by 0.32%, 1.62% and 1.62% respectively. This shows that using graph adjacency matrices Z_k and C_k at the same time can significantly improve the overall performance of the model, and verifies the optimization effect of the fusion graph adjacency matrix on the model performance. This is because the graph adjacency matrix C_k can dynamically adjust the connection strength between nodes by capturing the interaction between nodes and using the non-local attention mechanism, thus reflecting the complex characteristics in the movement process more accurately. At the same time, the introduction of C_k ensures that the model can comprehensively consider the interaction between joints when dealing with dynamic motion data. This fusion not only improves the accuracy of feature extraction, but also enhances the sensitivity of the model to subtle changes. To sum up, AGCM module significantly optimizes the overall performance of motion posture recognition by effectively integrating Z_k and C_k, which fully proves its unique contribution in improving the expression ability of the model.

Secondly, the performance of RCAM module is verified, and the result is shown in **Figure 7**.

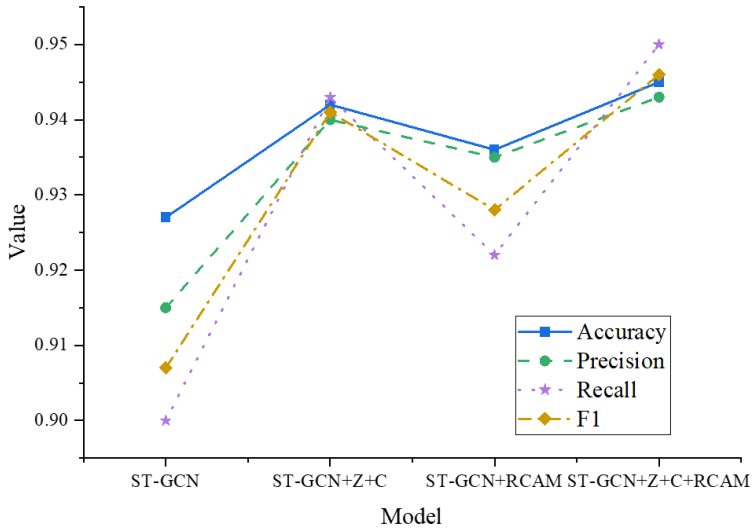

Figure 7. Experimental results of ablation based on RCAM module.

In **Figure 7**, the ST-GCN + Z + C + RCAM model, which combines the adjacency matrix of two graphs and RCAM module, performs best in all indicators, with an accuracy of 0.945, a precision of 0.943, a recall of 0.95 and a $F1$ value of 0.946. Compared with ST-GCN, the accuracy increased by 1.94% and the $F1$ value increased by 4.3%. Compared with ST-GCN + Z + C model without RCAM module, the accuracy is improved by 0.32% and the $F1$ value is improved by 0.53%. These data show that the introduction of RCAM module can improve the performance of the

model. This is mainly because RCAM module effectively enhances the model's ability to capture key information by weighting the importance of different characteristic channels. Specifically, RCAM module embeds the global information of each channel of the input feature map through global average pooling operation, and ensures that the important features are strengthened while the secondary features are suppressed through adaptive weight distribution. This feature enhancement mechanism enables ARST-GCN model to capture subtle changes in motion more accurately, especially in posture recognition tasks under complex motion patterns.

(2) Comparison with other algorithms

To better validate the effectiveness of the ABST-GCN model, comparisons are conducted on the NTU RGB + D dataset. The model is compared with Deep Long Short-Term Memory (D-LSTM), Variational Autoencoder Long Short-Term Memory (VA-LSTM), TCN, Adaptive Spatio-Temporal Graph Convolutional Network (AS-GCN), as well as the Spatio-Temporal Transformer (ST-TR) model and Pose Transformer (PoseFormer) model, which were both proposed in recent years specifically for human motion recognition tasks and feature innovative approaches in spatio-temporal feature modeling. The results are shown in **Figure 8**.

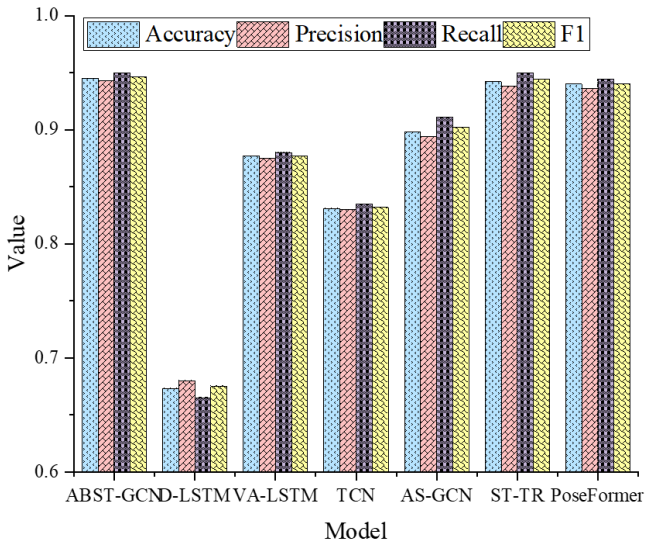

Figure 8. Performance comparison of different motion posture recognition algorithms.

In **Figure 8**, the ABST-GCN model performs well in the task of motion posture recognition, with its accuracy of 0.945, precision of 0.94, recall of 0.95 and *F*1 value of 0.945. Compared with the second performance ST-TR model, the accuracy, precision and *F*1 value are improved by 0.32%, 0.53% and 0.21% respectively. On the whole, ABST-GCN performs well in all performance indexes, which shows its remarkable advantages in dealing with complex motion posture recognition tasks.

(3) Comparison of model computing resources

In order to evaluate the feasibility and efficiency of ARST-GCN model in practical application, the differences in computing resources and training time between ARST-GCN and ST-GCN models under the same hardware environment are compared, and the results are shown in **Table 3**.

Table 3. Comparison of computing resources and training time between ARST-GCN and ST-GCN.

Model	Parameter quantity	Training time (hours)	GPU utilization (%)	Memory requirements (GB)
ST-GCN	3.1M	1.8	80	4
ARST-GCN	3.5M	2.7	85	6

Table 3 shows that ARST-GCN model has more parameters than ST-GCN in terms of the number of model parameters, mainly due to the introduction of AGCM and RCAM modules. This makes the total parameters of ARST-GCN model reach about 3.5M, while that of ST-GCN model is about 3.1M. Although the increase of parameters may improve the expressive ability of the model, it also leads to higher computational complexity. The training time of ARST-GCN model is slightly longer. Specifically, in the same hardware environment, the training time of ST-GCN model is 1.8 hours, while that of ARST-GCN model is 2.7 h. This difference is mainly because ARST-GCN needs to perform extra calculation steps in the training process to deal with the complexity introduced by AGCM and RCAM modules. Memory requirement is also an important index to evaluate the computing resources of the model. The memory requirement of ARST-GCN model is relatively high, about 6GB, while that of ST-GCN is 4GB. This means that using ARST-GCN model may face certain challenges in the environment of limited resources. Finally, the comparison of GPU utilization shows that the GPU utilization rate of ARST-GCN model is 85% during training, while that of ST-GCN model is 80%. This shows that ARST-GCN model can use GPU resources more effectively in the training process, but it also further increases the computational burden. On the whole, although ARST-GCN model has significantly improved its performance, the increase of its computing resources and training time has also brought challenges to practical application.

(4) Confusion matrix

The confusion matrix results of ARST-GCN model for five different action categories on NTU RGB+D dataset are shown in **Figure 9**.

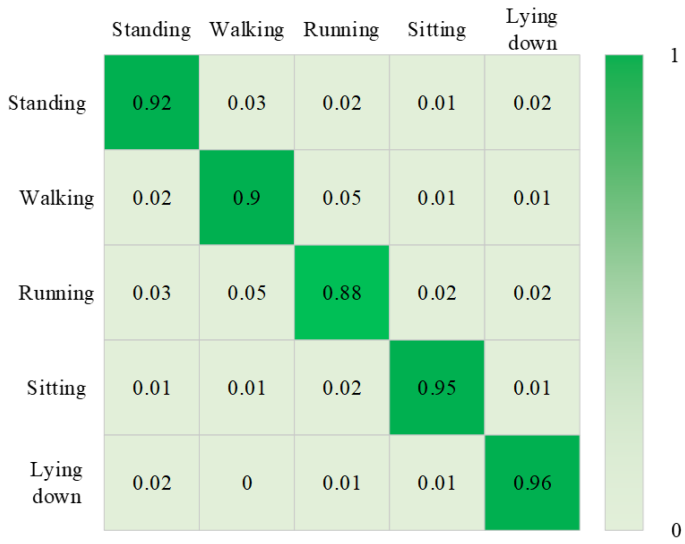

Figure 9. Accuracy confusion matrix of ARST-GCN model in five different actions.

In **Figure 9**, ARST-GCN model shows high recognition accuracy in most action categories, especially "lying" and "sitting" actions, and the accuracy in the confusion matrix is significantly higher than other values, indicating that these actions have better recognition effects. This is due to the obvious characteristics of these two actions in spatial posture. When lying down, the body is relatively static, and in most cases, the joint angle of the body changes little, which makes it easier for the model to extract relevant features. In the standing posture, although there is a certain dynamic, the overall action pattern is relatively stable and easy to identify. There is certain confusion between some action categories, such as "walking" and "running", which may be due to the biomechanical similarity between the two actions. Although running is faster than walking, and the gait is different, they may show similar skeletal movement patterns in the same environment and perspective. Therefore, ARST-GCN model may still need further optimization in capturing these dynamic features.

4.3. Performance verification of SIRA model

To confirm the effectiveness of the suggested model, the SIRA model utilizing XGBoost is contrasted with models utilizing Random Forest (RF), Decision Tree (DT), and Logistic Regressive (LR). In **Figure 10**, the outcome is displayed.

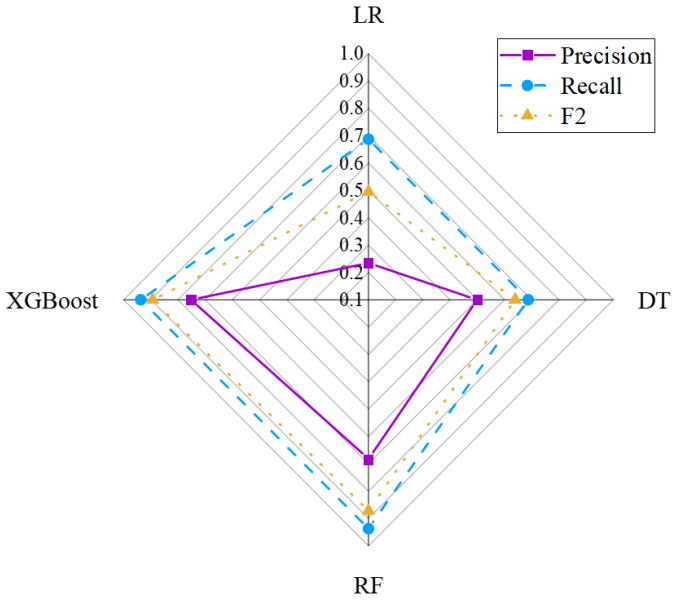

Figure 10. Performance comparison results of different SIR models.

Figure 10 shows that the precision, recall and *F*2 of XGBoost are 0.752, 0.937 and 0.893, respectively, and the overall performance is the most balanced and excellent, especially in the aspect of recall, which is very important for SIRA, because a higher recall means fewer real injuries are missed. This result shows that XGBoost has obvious advantages in dealing with SIRA tasks, and because of its higher recall rate and *F*2 value, it is especially suitable for application scenarios that require highly accurate detection.

5. Conclusion

By optimizing ST-GCN, this paper puts forward a SIRA model based on the optimized ARST-GCN, and evaluates the performance of this model through experiments. The conclusions are as follows: (1) In the ablation experiment based on AGCM, the model using graph adjacency matrix Z_k and C_k at the same time has the best accuracy of 0.942, precision of 0.94, recall of 0.943 and $F1$ value of 0.941, which verifies the optimization effect of the fused AGCM module on the model performance. (2) In the RCAM-based ablation experiment, ARST-GCN combined with AGCM and RCAM modules performed the best in all indexes, and compared with the model only combined with AGCM, the accuracy increased by 0.3% and the $F1$ value increased by 0.5%. This shows that the introduction of AGCM module and RCAM module significantly improves the performance of the model, which proves the effectiveness of RCAM module in improving the overall ability of motion posture recognition. (3) In the performance comparison of different SIR models, the precision, recall and $F2$ of XGBoost are 0.752, 0.937 and 0.893, respectively, and the overall performance is the most balanced and excellent, especially in the recall, which shows that XGBoost has obvious advantages in dealing with SIRA tasks.

However, there are still some shortcomings in this paper. The implementation of ARST-GCN model in real sports training also faces some practical challenges. Firstly, the model has high computational complexity. This complexity is particularly evident in practical application scenarios. In these scenarios, the model may need to process a large number of real-time data. This requirement places a high demand on hardware resources and computing power. How to effectively optimize the calculation efficiency and ensure the performance of the model is a big challenge in the process of model deployment. Secondly, the equipment for collecting motion data in sports training needs to have high accuracy and reliability, but the high-performance motion capture system is often expensive, which may limit the cost in large-scale implementation. In addition, the action characteristics of different sports events are quite different, and the adaptability of the model in different sports scenes still needs further verification. Future research can explore the generalization ability of the model in various sports environments, and how to reduce the dependence on expensive equipment and promote wider practical application. It can also be further extended to the combination of wearable technology to realize real-time sports data collection and risk assessment and explore the applicability of the model in different sports to verify its wide application potential. These directions will help to improve the practicality and popularization value of the model in actual sports training.

Author contributions: Conceptualization, ZH; methodology, ZH and YX; software, ZH; validation, ZH and YX; formal analysis, ZH and YX; investigation, ZH and YX; resources, ZH; data curation, YX; writing—original draft preparation, YX; writing—review and editing, ZH and YX; visualization, YX; supervision, ZH; project administration, ZH; funding acquisition, ZH and YX. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Fonti F, Ross J M, Aversa P. Using sports data to advance management research: A review and a guide for future studies. Journal of Management, 2023, 49(1): 325–362.
2. Qi Yufei, Sajadi S M, Baghaei S, et al. Digital technologies in sports: Opportunities, challenges, and strategies for safeguarding athlete wellbeing and competitive integrity in the digital era. Technology in Society, 2024: 102496.
3. Cui Jiayi, Du Hongyi, Wu Xiangzhi Data analysis of physical recovery and injury prevention in sports teaching based on wearable devices. Preventive medicine, 2023, 173: 107589.
4. Patricios J S, Schneider K J, Dvorak J, et al. Consensus statement on concussion in sport: the 6th International Conference on Concussion in Sport–Amsterdam, October 2022. British journal of sports medicine, 2023, 57(11): 695–711.
5. Ayala R E D, Granados D P, Gutiérrez C A G, et al. Novel Study for the Early Identification of Injury Risks in Athletes Using Machine Learning Techniques. Applied Sciences, 2024, 14(2): 570.
6. Guo Keyou, Wang Pengshuo, Shi Peipeng, et al. A New Partitioned Spatial–Temporal Graph Attention Convolution Network for Human Motion Recognition. Applied Sciences, 2023, 13(3): 1647.
7. Raj M S S, George S N, Raja K. Leveraging spatio-temporal features using graph neural networks for human activity recognition. Pattern Recognition, 2024, 150: 110301.
8. Meng Linsheng, Qiao Endong. Analysis and design of dual-feature fusion neural network for sports injury estimation model. Neural Computing and Applications, 2023, 35(20): 14627–14639.
9. Nassis G, Verhagen E, Brito J, et al. A review of machine learning applications in soccer with an emphasis on injury risk. Biology of sport, 2023, 40(1): 233–239.
10. Dhanke J A, Maurya R K, Navaneethan S, et al. Recurrent neural model to analyze the effect of physical training and treatment in relation to sports injuries. Computational Intelligence and Neuroscience, 2022, 2022(1): 1359714.
11. Rebelo A, Martinho D V, Valente-dos-Santos J, et al. From data to action: a scoping review of wearable technologies and biomechanical assessments informing injury prevention strategies in sport. BMC sports science, medicine and rehabilitation, 2023, 15(1): 169.
12. Li Ning, Zhu Xiaoyun. Design and application of blockchain and IoT-enabled sports injury rehabilitation monitoring system using neural network. Soft Computing, 2023, 27(16): 11815–11832.
13. Hu Kai, Ding Yiwu, Jin Junlan, et al. Skeleton motion recognition based on multi-scale deep spatio-temporal features. Applied Sciences, 2022, 12(3): 1028.
14. Wang Junping. Motion recognition based on deep learning and human joint points. Computational Intelligence and Neuroscience, 2022, 2022(1): 1826951.
15. Arab H, Ghaffari I, Chioukh L, et al. A convolutional neural network for human motion recognition and classification using a millimeter-wave Doppler radar. IEEE Sensors Journal, 2022, 22(5): 4494–4502.
16. Dong Zhe, Wang Xiongying. An improved deep neural network method for an athlete's human motion posture recognition. International Journal of Information and Communication Technology, 2023, 22(1): 45–59.
17. Li Nian, Boers S. Human Motion Recognition in Dance Video Images Based on Attitude Estimation. Wireless Communications and Mobile Computing, 2023, 2023(1): 4687465.
18. Adriaensen S, Biedenkapp A, Shala G, et al. Automated dynamic algorithm configuration. Journal of Artificial Intelligence Research, 2022, 75: 1633–1699.
19. Song Jae-Hun, Kong K, Kang S J. Dynamic hand gesture recognition using improved spatio-temporal graph convolutional network. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(9): 6227–6239.

20. Ali A, Zhu Yanmin, Zakarya M. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. Neural networks, 2022, 145: 233–247.
21. Li Maosen, Chen Siheng, Zhao Yangheng, et al. Multiscale spatio-temporal graph neural networks for 3d skeleton-based motion prediction. IEEE Transactions on Image Processing, 2021, 30: 7760–7775.
22. Lee B, Hong S, Kim H. Determination of workers' compliance to safety regulations using a spatio-temporal graph convolution network. Advanced Engineering Informatics, 2023, 56: 101942.
23. Spinelli I, Scardapane S, Uncini A. Adaptive propagation graph convolutional network. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(10): 4755–4760.
24. Kazi A, Cosmo L, Ahmadi S A, et al. Differentiable graph module (dgm) for graph convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1606–1617.
25. Cao Congqi, Zhang Xin, Zhang Shizhou, et al. Adaptive graph convolutional networks for weakly supervised anomaly detection in videos. IEEE Signal Processing Letters, 2022, 29: 2497–2501.
26. Mekruksavanich S, Jitpattanakul A, Sitthithakerngkiet K, et al. Resnet–se: Channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors. IEEE Access, 2022, 10: 51142–51154.
27. Park K, Soh J W, Cho N I. A dynamic residual self-attention network for lightweight single image super-resolution. IEEE Transactions on Multimedia, 2021, 25: 907–918.
28. Lövdal S S, Den Hartigh R J R, Azzopardi G. Injury prediction in competitive runners with machine learning. International journal of sports physiology and performance, 2021, 16(10): 1522–1531.