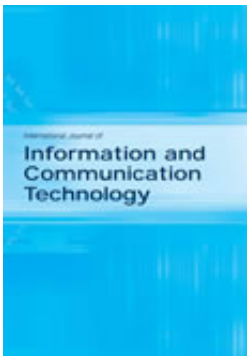


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Sports movement analysis method considering early fusion network structure and key points of human body

Jianjun Yin, Jie Chen

DOI: 10.1504/IJICT.2025.10074850

Article History:

Received:	08 August 2025
Last revised:	22 October 2025
Accepted:	24 October 2025
Published online:	12 December 2025

Sports movement analysis method considering early fusion network structure and key points of human body

Jianjun Yin

Department of Physical Education,
Guangdong University of Finance and Economics,
Guangzhou, 510320, China
Email: dajun8489@163.com

Jie Chen*

School of Humanities and Communication,
Guangdong University of Finance and Economics,
Guangzhou, 510320, China
Email: jiechenjcj@outlook.com
*Corresponding author

Abstract: With the rapid development of sports and computer technology, accurate sports movement analysis has become crucial for enhancing athlete performance and rehabilitation. Traditional methods face challenges such as difficulty in recognising multi-scene actions and inconsistent sequence lengths. To address this, a novel approach combining an early fusion network with human key point data is proposed. By integrating skeleton node information and using Neville interpolation, the method enhances feature extraction and temporal localisation. Experimental results show significant improvements: compared to traditional models such as LSTM and ST-GCN, the EF-GCN model proposed in this study achieves an increase in classification accuracy of up to 18.5% across various neural networks, and performance metrics such as accuracy, precision, recall, and F1-score improve by around 10%. This approach offers substantial advancements in motion analysis and holds great potential for future sports training and rehabilitation applications.

Keywords: early fusion network structure; key points of the human body; sports movement analysis; Neville interpolation method; temporal positioning.

Reference to this paper should be made as follows: Yin, J. and Chen, J. (2025) 'Sports movement analysis method considering early fusion network structure and key points of human body', *Int. J. Information and Communication Technology*, Vol. 26, No. 43, pp.36–60.

Biographical notes: Jianjun Yin obtained his Master's degree from South China Normal University. He is currently working at Guangdong University of Finance and Economics. His main research direction is social sports guidance.

Jie Chen holds a Master's degree from Wuhan University. He is currently working at Guangdong University of Finance and Economics. His main research direction is news publishing.

1 Introduction

With the popularisation and continuous improvement of sports competition, the demand for sports technical movement analysis is increasing day by day. Sports movements are not only directly related to the performance and level of competition of athletes, but also important factors that affect the scientificity, efficiency, and other aspects of sports training. Sports movement analysis can help athletes optimise their technical movements, which is more conducive to the prevention and rehabilitation of sports injuries. Therefore, sports movement analysis is extremely important for athletes. In view of this, domestic and foreign researchers have conducted extensive research on it. For example, Yu and Bo proposed a method that combines background subtraction algorithm with pedestrian detection technology to address the impact of sports model body movements and clothing prints on online advertising effectiveness. The outcomes revealed that this method could validly improve detection rate and accuracy, and scientifically and reasonably evaluate advertising effectiveness (Yu and Bo, 2024). Geisen et al. proposed a motion difference recognition method based on sensor clothing data to address the limitations of existing motion classification schemes. The results indicated that this method was feasible in sports training. By comparing the data of volleyball serving and handball standing throwing, it can be found that there were subtle differences in different operating techniques of the same athlete, which helped to deepen the understanding of the operating factors that affect specific sports movements (Geisen et al., 2024). Malawski proposed a method for real-time qualitative action analysis that combines depth and inertial sensor data to address the issue of providing useful feedback in sports. The outcomes revealed that this method could validly detect and analyse fencing stabbing actions, and the accuracy and efficiency of depth and inertia sensors in analysing fencing steps were verified through comparative experiments (Malawski, 2021). Fan and Lin proposed an automatic recognition method based on thermal imaging and deep learning to address the issue of how to use intelligent infrastructure to monitor athlete sports injuries. The outcomes revealed that this method could validly identify lower limb sports injuries, with an average error of less than 2.22% in detecting the severity of injuries, and the output results were better than traditional methods (Fan and Lin, 2025). Miyake and Miyake proposed a method that combines optical sensors and six axis inertial sensors to address the limitations of traditional muscle deformation analysis methods in finger force estimation. The results indicated that this method could validly improve the accuracy of finger force estimation (Miyake and Miyake, 2025). Ma et al. proposed an unsupervised action segmentation framework based on motion principles to address the issues of insufficient fine-grained analysis in action segmentation and lack of interpretability in motion representation. The outcomes revealed that the method exhibited good performance and universality in different subjects, datasets, and application scenarios (Ma et al., 2022).

The critical importance of sports movement analysis stems from its direct and profound impact on enhancing athletic performance and safeguarding athletes' health. In elite competitive sports, subtle optimisations in technical movements often determine the outcome of a competition. Through precise movement analysis, coaches can provide data-driven feedback to athletes, enabling the optimisation of movement efficiency, power output, and energy distribution, thereby facilitating scientific training. Furthermore, this technology plays a pivotal role in the prevention and rehabilitation of

sports injuries. By identifying non-standard movement patterns that may lead to injuries (such as improper landing postures or asymmetrical gaits), proactive intervention and correction can be implemented. For injured athletes, quantitative analysis of movement quality is indispensable for monitoring rehabilitation progress and ensuring a safe return to competition. However, numerous challenges remain in translating these urgent application needs into robust computational analysis solutions.

Although existing sports movement analysis methods have made significant progress, there are still issues with insufficient adaptability to dynamic features and deficiencies in data quality and integrity. With the ongoing advancement of computer technology, deep learning technology has gradually demonstrated superior effects in various fields. The method based on convolutional neural networks (CNN) provides a foundation for precise localisation of human keypoints, and early fusion of network structures as a key strategy to improve detection performance has become a hot research topic today. Hasanvand et al. proposed a new similarity analysis method based on a deep learning framework to extract confidence scores for human keypoints, in response to the need to improve the accuracy of human motion posture assessment in the fields of sports, dance, and healthcare. The outcomes revealed that this approach performed better than existing techniques on human pose image datasets and various image sets, and could more accurately evaluate human pose similarity through keypoint confidence scores (Hasanvand et al., 2023). Li et al. proposed a novel unsupervised human keypoint detection scheme to address the issue of unreasonable keypoint allocation caused by reconstructed images in unsupervised human keypoint detection. The outcomes revealed that the performance of this approach was superior to existing methods on multiple popular datasets. On the dataset, the effective version improved performance by 7.0%, and the efficient version improved performance by 5.7% without sacrificing inference speed (Li et al., 2025). Zhang et al. (2021) proposed an efficient network structure, three effective training strategies, and four useful post-processing techniques to address the challenges faced by single image human keypoint detection, such as occlusion, blur, lighting, and scale changes. The results indicated that this approach outperformed representative state-of-the-art methods in keypoint detection benchmark tests (Zhang et al., 2021). Zamani and Baleghi proposed an early fusion structure method based on visible light and thermal imaging fusion for weed detection in rice fields. The results indicated that this approach effectively improved the accuracy of weed detection by combining the features of two image modalities (Zamani and Baleghi, 2023). Priyanka and Kumar proposed a novel early and late fusion CNN for multi-channel speech enhancement problems. The results indicated that these models performed better than existing popular methods on multi-condition microphone array data (Priyanka and Kumar, 2023).

In summary, existing research has achieved excellent results in sports action analysis methods, early fusion structures, and key points of the human body. However, most methods have problems such as difficulty in distinguishing and recognising targets and actions in multi-person scenes, strong perspective dependence, and inconsistent action sequence lengths. Therefore, a new method for sports action analysis combining early fusion network structure and human key points is proposed to overcome the above problems and improve the accuracy of action classification and temporal positioning. The innovation of the research lies in the proposal of early fusion graph convolutional network (EF-GCN), which adopts the Neville interpolation method and smooth frame discrimination module to improve the accuracy of action timing localisation and the

completeness of feature extraction, providing a more efficient and accurate method for sports action analysis.

2 Methods and materials

The study first constructs a sports action classification model based on EF-GCN. Secondly, a sports action localisation model is constructed by combining key points of the human body, which adopts the Neville interpolation method and a smooth frame discrimination module optimisation algorithm. Finally, a sports action analysis model based on early fusion network structure and key points of the human body is constructed.

2.1 Sports action classification model based on EF-GCN

The classification of sports movements is the foundation and core prerequisite for sports analysis. Due to differences in rules, techniques, and physical fitness requirements among different sports, accurate classification is necessary for targeted and in-depth analysis. In the field of sports action classification, existing methods have achieved superior results in single target recognition and fixed angle action capture (Shakrani et al., 2022; Di Domizio and Fabrizi, 2024). However, there are still shortcomings in distinguishing and recognising target individuals and actions in multi-person scenes, as well as reducing perspective dependence. EF-GCN can fuse multi-modal data in the early stages to enhance the integrity of feature extraction and improve action classification accuracy (Su et al., 2025; Sasikaladevi et al., 2024). Therefore, a sports action classification model based on EF-GCN is proposed, which effectively integrates the node and edge information of the human skeleton to enhance the accuracy of action classification. The sports action classification model first fuses the position information (nodes) of human key points and the direction and length information (edges) of bones in the early fusion layer, as shown in equation (1).

$$F = Fuse(V, E)\begin{cases} V \in R^{N\times C} \\ E \in R^{M\times C} \end{cases} \tag{1}$$

In equation (1), V represents the node information matrix. E represents the edge information matrix. M indicates the number of edges. C represents the coordinate dimension. N indicates the number of key points. *Fuse* indicates the method of concatenation or weighted fusion. This formula model can fully consider the structural information of the human skeleton, fuse the features of nodes and edges together, and provide a richer data foundation for the action classification of the model. The study used a time-space dual stream feature extraction network, as shown in Figure 1.

Figure 1 shows the feature extraction process, which mainly includes two branches: spatial and temporal feature extraction. The input data is first divided into two parts, which are used for spatial and temporal feature extraction (Luo et al., 2024; Yu et al., 2024). The input data is processed through multiple layers and spatial features are extracted through spatial convolution. In the temporal features, multiple layers of processing are also applied to extract temporal features. Finally, the feature fusion module combines the extracted spatial and temporal features to obtain fused features. The graph convolutional layer updates node features through aggregation operations, which

involves aggregating information from nodes and their neighbouring nodes to update node feature representations, thereby capturing the spatial structural relationships of the human skeleton (Bao et al., 2025). The graph convolution operation is shown in equation (2).

$$h_i' = \sigma\left(\sum_{j \in N(i)} W \cdot h_j + b\right) \tag{2}$$

In equation (2), $N(i)$ represents the set of neighbouring nodes of node v_i. W represents the weight matrix. h_j represents the eigenvectors of neighbouring node j. b indicates the bias term. σ represents the activation function. h_i' represents the updated feature vector of node i. Through graph convolution operations, the model can effectively learn the interrelationships between key points in the human body, extract discriminative spatiotemporal features, and provide key feature support for the classification of sports movements (Mishina et al., 1995). The schematic representation of the early fusion network is shown in Figure 2.

Figure 1 Schematic diagram of dual-stream network (see online version for colours)

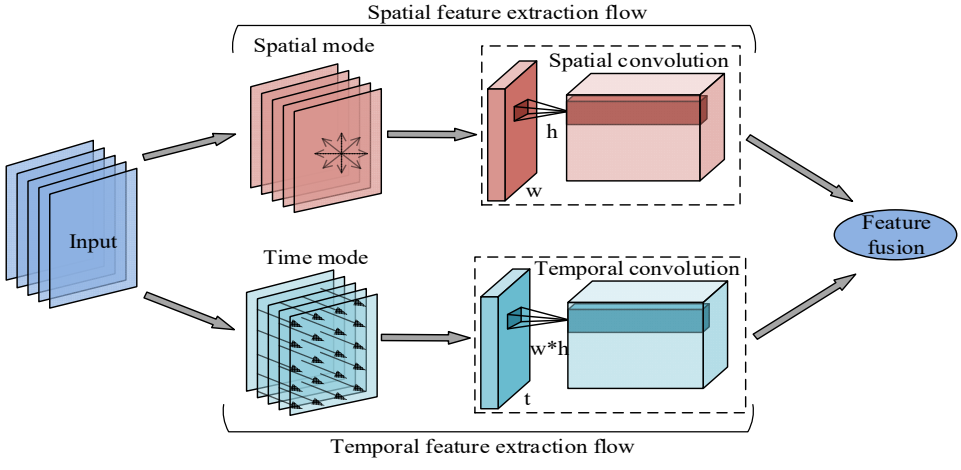


Figure 2 Schematic representation of the early fusion network (see online version for colours)

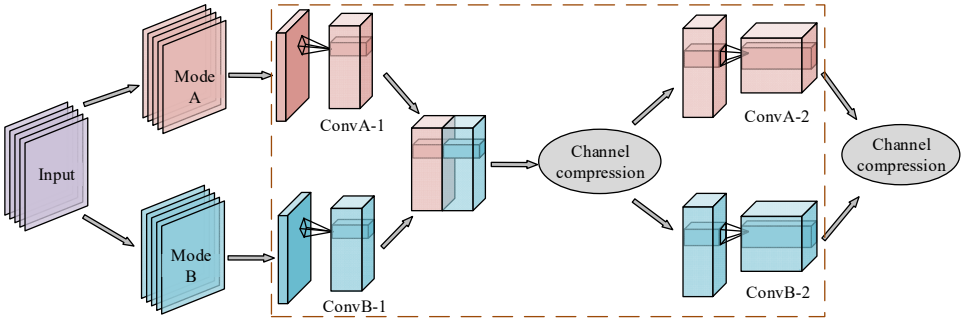


Figure 2 presents a schematic representation of an early-fusion network. Initially, the input data is divided into temporal and spatial modality branches (such as RGB videos and depth maps) for processing. Both modalities undergo convolution operations to extract preliminary features. Subsequently, the processed data is concatenated and subjected to channel compression. This step removes redundant information while highlighting important features. Afterward, through a feature interaction structure, features from different modalities complement and fuse with each other, enabling the network to comprehensively utilise multi-modal information. The fused features then undergo convolution operations again to further extract higher-level features. This process demonstrates that the early-fusion network integrates data from different modalities at an early stage, enabling the network to exhibit stronger capabilities in subsequent feature extraction and task execution. The dynamic topology layer updates the connection relationships between nodes in real-time based on action states to adapt to the variations and requirements of different sports actions. Specifically, this is illustrated as shown in equation (3).

$$A_{dynamic} = A + P \begin{cases} A \in R^{O \times O} \\ P \in R^{O \times O} \end{cases} \tag{3}$$

In equation (3), $A \in R^{O \times O}$ represents the basic adjacency matrix, which expresses the static connection relationship of the human skeleton. $P \in R^{O \times O}$ represents a dynamic weight matrix. O represents the total number of nodes in the human skeleton. $A_{dynamic}$ represents a dynamic adjacency matrix. Based on this dynamic topology structure, the model can adjust the connection relationship of the human skeleton in real-time according to the actual motion pattern of sports movements, thereby more accurately capturing the spatiotemporal characteristics of movements and improving the motion classification performance of the model (Endo et al., 2023). To further highlight the importance of key nodes in sports action classification, the model assigns different weights to key points. The specific implementation is shown in equation (4).

$$agg(V) = V \odot P \tag{4}$$

In equation (4), $\odot$ represents element wise multiplication. In this way, the model can enhance the attention of nodes that play a major role in sports, allowing these key nodes to play a greater role in feature extraction and analysis, thereby improving the model's ability to recognise and classify sports movements. The propagation direction of features is crucial for the effectiveness of feature extraction, with bidirectional propagation being the most effective (Dornaika, 2023). It can more comprehensively capture the motion characteristics of the human body in sports movements, thereby extracting richer spatiotemporal information. The schematic diagram of one-way graph propagation and two-way graph propagation pathways is shown in Figure 3.

Figure 3 shows the propagation paths of unidirectional graph and bidirectional graph. Figure 3(a) is a unidirectional graph propagation pathway characterised by the one-way transmission of information from one node to another. Figure 3(b) shows a bidirectional propagation path, where information not only propagates in the forward direction, but also has feedback or influence in the reverse direction. In most sports movements, the support points, force points, and focus points of the human body will change, and the bidirectional graph structure can provide each element in the graph with a global view.

This approach can improve the accuracy of sports action classification. This enables bidirectional communication to exhibit better feature extraction performance in sports action analysis, enabling more accurate recognition and classification of various sports actions. Using EF-GCN to extract spatiotemporal features of human keypoints, this process transforms the complex motion patterns of human keypoints in sports movements into representative feature representations. As shown in equation (5).

$$H = Extract(F) \tag{5}$$

In equation (5), H represents the extracted feature representation. This process provides key feature support for subsequent classification tasks, enabling classifiers to accurately classify sports actions based on these features. Based on the extracted features, the research designs an action classifier to classify sports actions, as shown in equation (6).

$$\begin{cases} S = W_{fc} \cdot H + b_{fc} \\ H \in R^{N \times D} \end{cases} \tag{6}$$

In equation (6), D represents the feature dimension. W_{fc} represents the weight matrix of the fully connected layer. b_{fc} indicates the bias term. Function W_{fc} is used to convert the score into a probability distribution, as shown in equation (7).

$$P = Softmax(S) \tag{7}$$

In equation (7), P represents the output probability distribution vector. In this way, the classifier can calculate the probability of each action category based on the extracted features, thereby achieving the classification of sports actions. To measure the difference between the predicted probability distribution and the true labels and guide the training process of the model, a cross entropy loss function (LF) is used, as shown in equation (8).

$$L_{cls} = -\sum_{i=1}^{C} y_i \log P_i \tag{8}$$

In equation (8), y_i represents the true label. P_i represents the predicted probability. By minimising the LF, the model can continuously optimise its own parameters, improve the accuracy of sports action classification, and make the predicted results as close as possible to the true labels (Xiao et al., 2023). In summary, the research has completed the construction of a sports action classification model based on EF-GCN, and its overall flowchart is shown in Figure 4.

In Figure 4, after RGB video input, data pre-processing is performed first, and a sequence of key points is generated to construct the skeleton map. After integrating multi-modal information in the early fusion layer, feature extraction is performed through spatial feature branches (skeleton graph convolution) and temporal feature branches (node graph convolution and temporal convolution). Next, the features are processed using dynamic topology layers and keypoint weight distributions, and their representation is enhanced through bidirectional graph propagation. Finally, the integrated features from the feature fusion module are fed into the action classifier, which uses softmax probability transformation and cross entropy loss for action classification and outputs the final action classification result. Through this flowchart, it is possible to construct an efficient and high-precision sports action classification model.

Figure 3 Schematic diagram of unidirectional graph propagation and bidirectional graph propagation paths, (a) the information dissemination channels of unidirectional graphs (b) bidirectional graph information dissemination channels (see online version for colours)

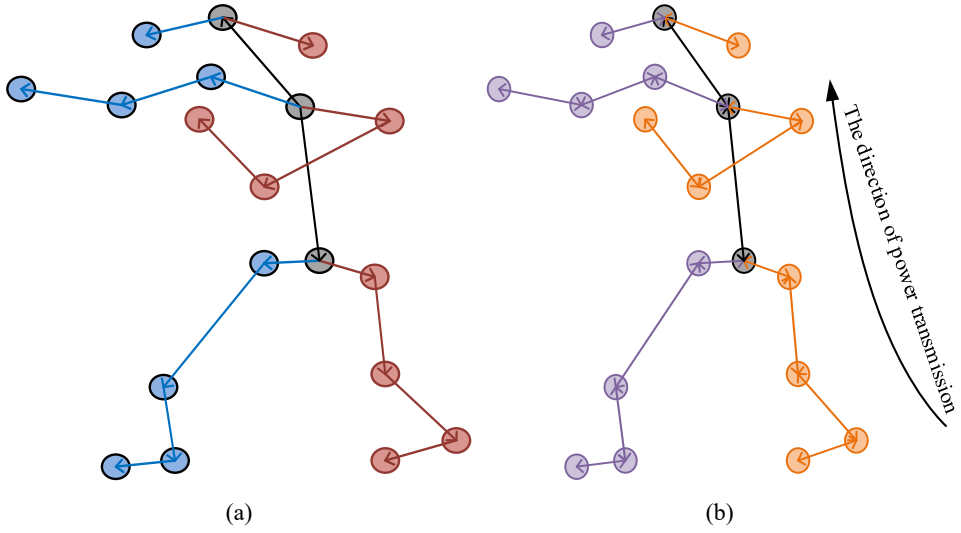
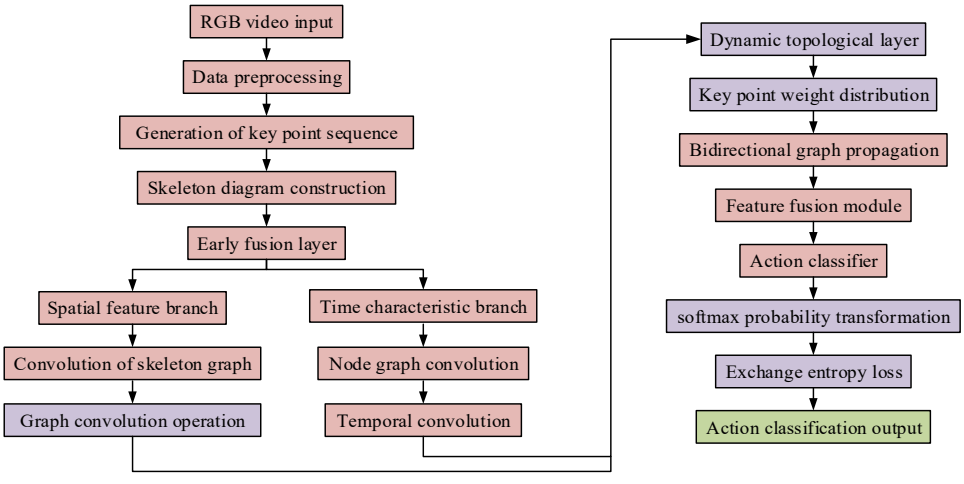


Figure 4 Overall flowchart of the sports action classification model based on EF-GCN (see online version for colours)



2.2 Sports movement analysis method combining EF-GCN and key points of the human body

In the previous section, a sports action classification model based on EF-GCN was studied and constructed, which achieved excellent results in integrating human skeleton nodes and edge information, effectively improving the accuracy of sports action classification. However, the model lacks adaptability to dynamic features and has limited

ability in temporal localisation, making it difficult to accurately determine the starting and ending time points of actions (Zhang et al., 2024). Therefore, the study introduces the graph convolutional network-based boundary-matching network (GCN-BMN). By introducing new data processing techniques and capturing and utilising the spatiotemporal features of key points in the human body, the classification and temporal localisation performance of sports technology movements can be improved, providing more accurate solutions for sports training and movement analysis. The requirements for human keypoint sequences in temporal localisation and action classification tasks are different, and it is necessary to ensure that effective data covers the complete dimensions of the input tensor. The comparison chart of action and timing positioning data is shown in Figure 5.

Figure 5 A comparison chart of action classification and temporal positioning data, (a) action classification sequence (b) temporal positioning sequence (see online version for colours)

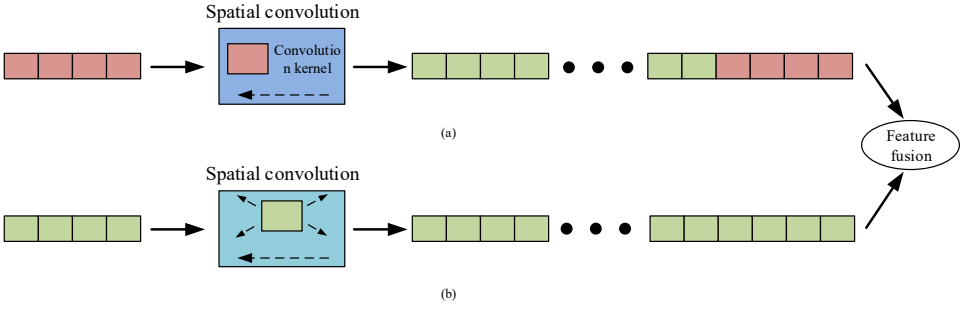


Figure 5(a) shows the spatial feature extraction under the action classification sequence. The input data undergoes spatial convolution operation, and the convolution kernel is used to extract features. The output feature map is used for subsequent action classification. Figure 5(b) displays spatial feature extraction under temporal localisation sequence, where the input data is also processed through spatial convolution, and the output feature map is used for temporal localisation. After extracting the spatial features of the two sequences, further feature fusion can be performed to provide richer information for subsequent action classification. In response to the problems of uneven quality and missing keyframes in sports action data, the Neville interpolation method was used to improve data augmentation, as shown in equation (9).

$$T_{i,j}^{(k)} = \begin{cases} T_{i,j-1}^{(k-1)} + \dfrac{\left(x - x_{i-j+1}\right)}{\left(x_i - x_{i-j+1}\right)}\left(T_{i,j-1}^{(k-1)} - T_{i-1,j-1}^{(k-1)}\right) & \text{if } i > j \\ T_{i,j}^{(k-1)} & \text{if } i = j \end{cases} \tag{9}$$

Through this interpolation method, missing data can be effectively filled, enhancing the integrity and availability of the data, and providing a more reliable data foundation for subsequent analysis (de Camargo, 2022). The collected human keypoint data are re-sampled to a fixed length that meets the model input requirements, as shown in equation (10).

$$S_{out} = Resample(S_{in}, L) \tag{10}$$

In equation (10), S_{in} represents the input signal or data sequence. L represents re-sampling parameters. *Resample* represents the re-sampling function. Data standardisation is a key step in ensuring data consistency (Li et al., 2024). Re-sampling data to a fixed length enables the model to stably receive and process input data, thereby ensuring the reliability and consistency of analysis results. The process expands the data in the time dimension and generates additional keypoint data using the Neville interpolation method to increase data diversity, as shown in equation (11).

$$S_{aug} = Augment(S_{in}, r) \tag{11}$$

In equation (11), r represents the enhancement parameter, which controls the degree or manner of enhancement. *Augment* represents the data augmentation function. Data augmentation can enrich the training dataset, expose the model to more diverse motion patterns, thereby improving the model's generalisation ability and adaptability to various sports movements (Mumuni and Mumuni, 2025). This is particularly important in sports movement analysis, as different athletes may perform the same movement in different ways. The key point data are smoothed using the sliding average method to reduce noise interference, as shown in equation (12).

$$S_{smooth} = Smooth(S_{in}, w) \tag{12}$$

In equation (12), w represents the smoothing parameter. *Smooth* represents a smoothing processing function. Smoothing processing helps to remove random noise from data, making the motion trajectories of key points in the human body smoother and more natural, and more in line with the characteristics of actual sports movements, thereby improving the analysis performance of the model (Movassagh et al., 2023). Temporal localisation is used in sports action localisation, and its temporal feature extraction is shown in equation (13).

$$F_{temp} = Conv1D(S_{in}, K, D) \tag{13}$$

In equation (13), K represents the size of the convolution kernel. D represents the stride of the convolution. $Conv1D$ represents one-dimensional convolution operation. Multi-scale feature extraction can fully capture the temporal variation patterns of sports movements, providing rich feature information for temporal localisation (Xue et al., 2023; Li et al., 2023). In this way, the model can more accurately identify the starting and ending time points of sports actions, thereby achieving precise positioning of action timing. Continuous actions can easily lead to sticking in time sequence. A smooth frame discrimination module has been introduced in the study to assign membership values to each moment, in order to more accurately determine the key time points of sports actions (Ma et al., 2024). There are usually four methods for initialising membership degrees, as shown in Figure 6.

Figure 6 Comparison of four membership degree initialisation functions, (a) one-time function initialisation (b) initialisation of the arctangent function (c) initialisation of the normal distribution function (d) cosine function initialisation (see online version for colours)

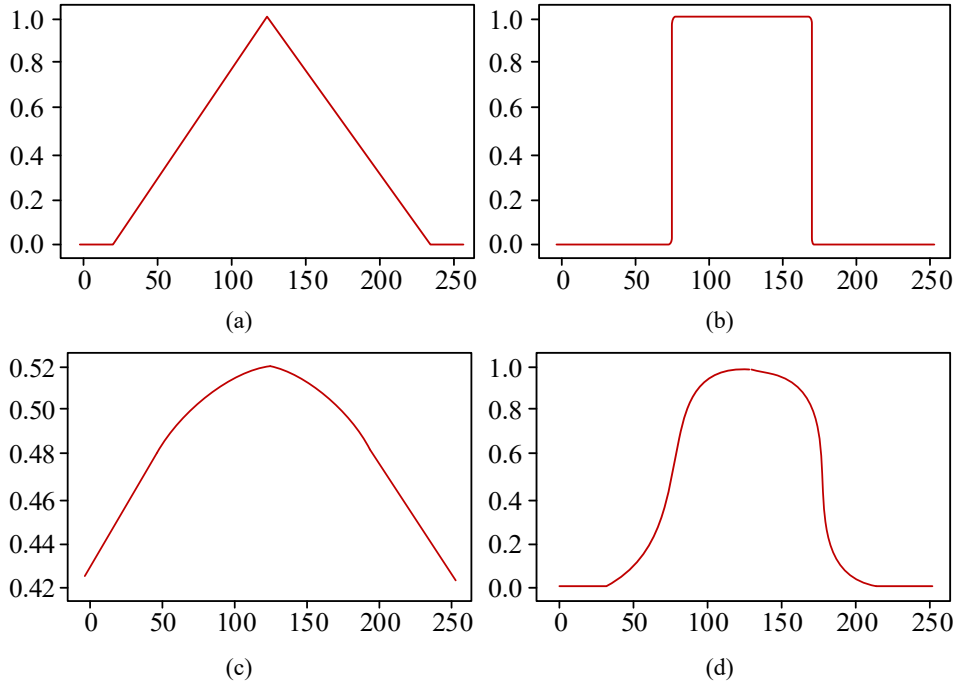


Figure 6(a) shows the initialisation of a linear function, which is too steep at the extremum, making the transition process not smooth enough. Figure 6(b) shows the initialisation of the normal distribution function. Although it can ensure smoothness, the membership degree before and after the action is not significantly different, making learning more difficult. Figure 6(c) shows the initialisation of the arctangent function, which avoids the problems of the above two functions. However, there is no point where the slope is 0, and there may be slight abrupt changes. Figure 6(d) shows the initialisation of the cosine function, and its energy curve can meet the requirements. Therefore, the study uses the cosine function to initialise the membership curve, as shown in equation (14).

$$f(x)=\begin{cases}0.5\cos\left(\dfrac{\pi(x-\mu)}{half}\right)+0.50 & \text{if } \mu-half\le x\le\mu+half\\ 0 & \text{otherwise}\end{cases} \tag{14}$$

In equation (14), μ represents the centre position parameter. *half* represents the half width parameter. x represents input variables. The shape of the membership curve is adjusted through optimisation algorithms to fit the true temporal structure of the action, as shown in equation (15).

$$\hat{f}(x)=Optimise\left(f(x)\right) \tag{15}$$

In equation (15), *Optimise* represents the optimisation function. The LF of the temporal localisation task comprehensively considers three aspects: boundary matching, boundary regression, and frame discrimination, to ensure that the model can accurately locate the position of sports actions in the time series. The formula is shown in equation (16).

$$L = \lambda_1 L_{temp} + \lambda_2 L_{reg} + \lambda_3 L_{seq} \tag{16}$$

In equation (16), L_{temp} represents the boundary matching loss. L_{reg} represents boundary regression loss. L_{seq} represents frame discrimination loss. λ_1, λ_2, and λ_3 represent the weight coefficients. By integrating these three types of losses, the model can comprehensively consider various aspects of temporal localisation tasks, thereby achieving accurate localisation of sports action timing. Based on the above model construction, the GCN-BMN structure is shown in Figure 7.

Figure 7 Full-scale deep supervision network architecture diagram (see online version for colours)

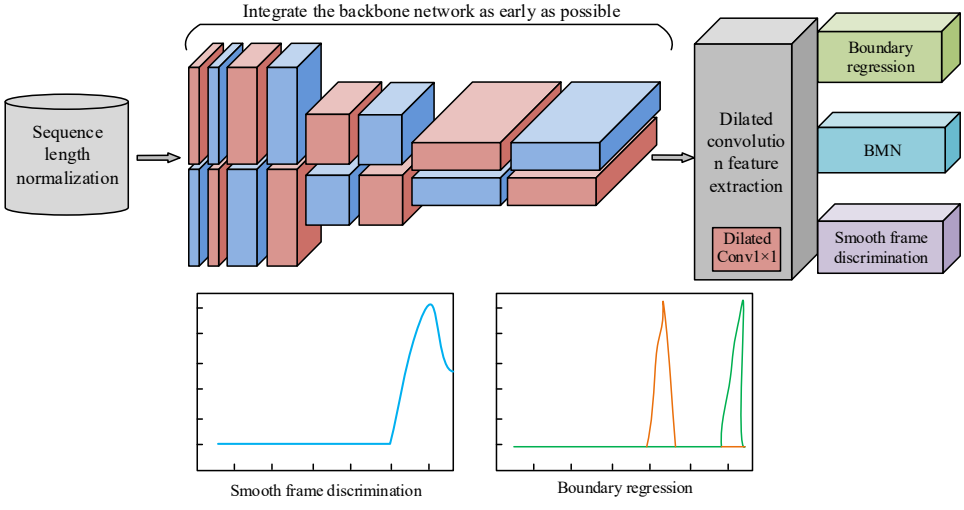
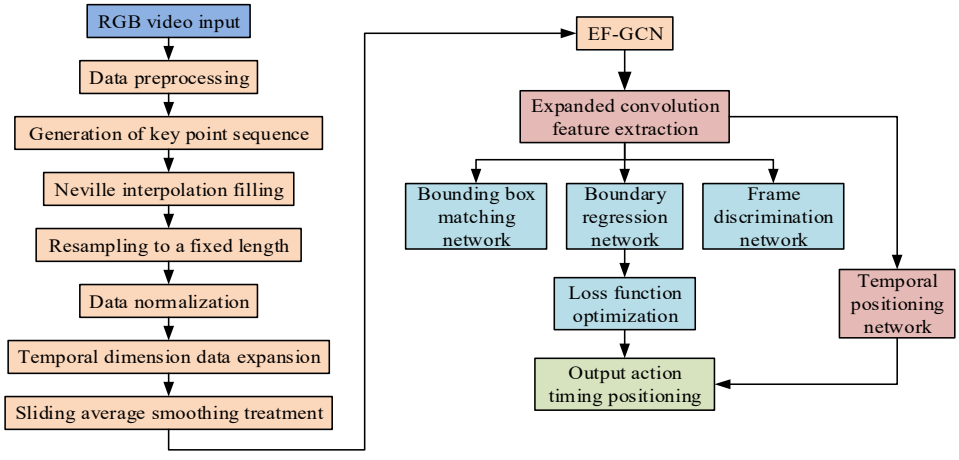


Figure 7 shows the GCN-BMN network structure diagram. Firstly, in the data pre-processing stage, the sequence length is normalised to generate sequence information with uniform length and rich data. Then, it is processed by a backbone network similar to early fusion. Then the net work enters the dilated convolution for feature extraction. The final extracted features are processed through boundary regression, boundary matching network, and smooth frame discrimination to output recognition results. In summary, the research has completed the construction of the early fusion and keypoint-based sports action analysis (EFK-SAA) model. The flowchart is shown in Figure 8.

Figure 8 shows the flowchart of the EFK-SAA model. Firstly, there is the pre-processing part of the video data, followed by the introduction of Neville interpolation filling, re-sampling, data normalisation, time dimension expansion, and smoothing processing. Afterwards, EF-GCN is used for feature extraction, followed by further processing of features through modules such as extended convolutional feature extraction, bounding box matching network, boundary regression network, and frame discrimination network. Finally, by optimising the LF, the model outputs the temporal

localisation result of the action. The entire process has achieved complete processing from video input to action classification and time localisation.

Figure 8 Flowchart of the EFK-SAA model (see online version for colours)



3 Results

The experiment first compared the performance indicators to verify the performance of EF-GCN, and then verified and analysed the performance of GCN-BMN. Finally, the performance loss and indicators of the overall model were compared and verified.

3.1 Verification of sports action classification method based on EF-GCN

The study first conducted experimental verification on the sports action classification performance of EF-GCN. Before the experiment, the research needed to prepare the required experimental environment and data. This included the hardware and software configurations required for the experiment, as well as the dataset for the experiment. Efficient hardware configuration and stable software version were sufficient to support the smooth progress of the experiment, with specific parameters shown in Table 1.

According to the parameters shown in Table 1, two datasets, Sports-1M and UCF-101, were introduced for experimental data preparation. The UCF-101 dataset has a rich and diverse range of action categories, with videos sourced from BBC/ESPN radio and television channels as well as video websites such as YouTube, demonstrating high diversity and representativeness. The UCF-101 dataset has a huge amount of data, including videos of various sports activities, which can provide rich data support for sports action analysis. In action classification detection, accuracy is often an important indicator for evaluating performance. For uneven sample distribution of data, accuracy indicators can alleviate this deficiency. Long short-term memory (LSTM) has wide applications in tasks such as sequence prediction and natural language processing. Spatio temporal graph convolutional networks (ST-GCN) is a neural network used for processing spatiotemporal sequences on graph structured data. Dynamic graph neural networks (DGNN) are suitable for processing dynamic graph data. The study compared

and validated the accuracy and precision of these different models on two datasets, as shown in Figure 9.

Table 1 Hardware and software configuration parameters table

Category	*Item*	*Model/version*
Hardware	Computer host	Intel Core i7-10700K, 16 GB DDR4 3,200 MHz RAM, 256 GB SSD, 1 TB HDD
	GPU accelerator card	NVIDIA RTX 3080, 10 GB+ VRAM
	Camera	1,920 × 1,080 resolution, 30fps frame rate, autofocus
	Image acquisition card	Compatible with camera, supports multiple resolutions and formats
	Electronic component stage	Adjustable angle and position, with lighting equipment
	Server	Multi-GPU configuration, scalable based on data and computation needs
Software	Operating system	Linux Ubuntu 20.04
	Deep learning framework	PyTorch 1.9. X
	Data processing tools	Python 3.8+, NumPy, Pandas, OpenCV
	Image annotation tool	labelImg
	Model evaluation tools	Scikit-learn, TensorBoard
	Visualisation tools	Matplotlib, Seaborn
	Database management system	MySQL
Experimental configuration	Optimiser	Adam
	Base learning rate	0.001
	Batch size	32
	Learning rate scheduler	StepLR (step_size = 30, gamma = 0.1)
	Training epochs	100
	Weight decay	0.0001
	Loss function	Cross-entropy (classification), combination of boundary matching, regression, and frame discrimination losses (localisation)

Figures 9(a) and 9(b) compare the classification accuracy and precision of different models on the Sports-1M and UCF-101 datasets. In the Sports-1M dataset, the proposed model improved accuracy by 18.5%, 1.5%, and 3.4% respectively compared to LSTM, ST-GCN, and DGNN models, and improved accuracy by 18.0%, 1.4%, and 3.3% respectively. On the UCF-101 dataset, the accuracy improved by 17.0%, 1.4%, and 2.6% respectively, and the accuracy improved by 16.4%, 1.5%, and 2.5% respectively. This indicated that the research model had a significant effect on improving the accuracy of sports action classification tasks. Training loss is an important indicator for measuring the fitting effect. The experiment further verified the training performance of the model by comparing this indicator, as shown in Figure 10.

Figure 9 Comparison of classification accuracy rates of different models on the Sports-1M and UCF-101 datasets, (a) comparison of classification accuracy rates in the Sports-1M dataset (b) comparison of classification accuracy rates in the UCF-101 dataset (see online version for colours)

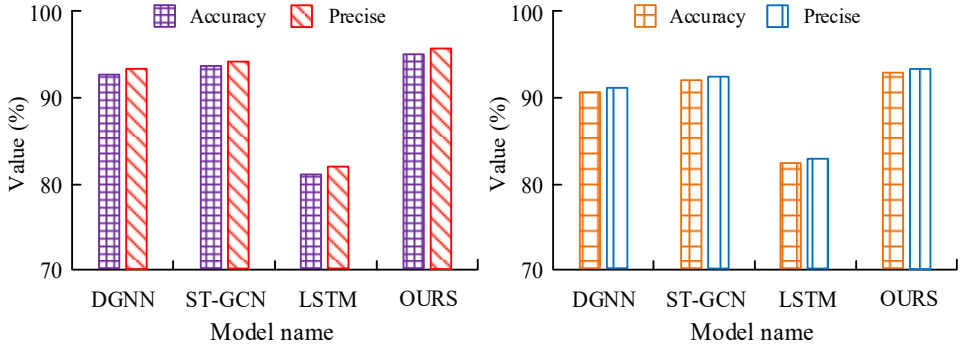
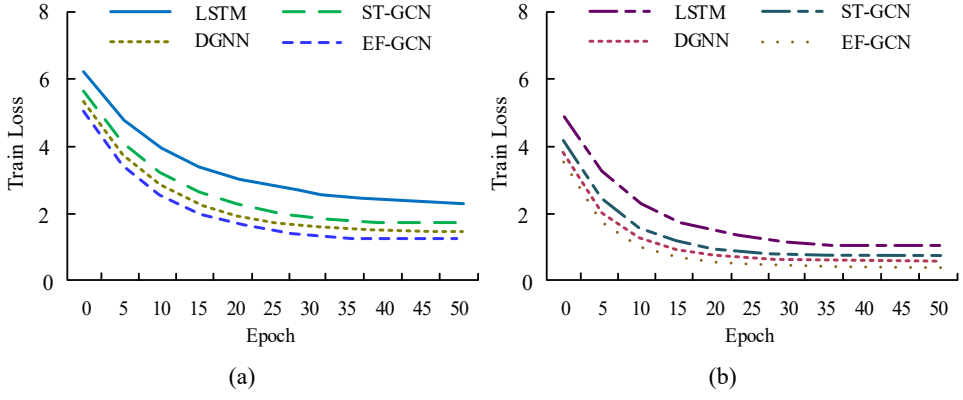


Figure 10 The training loss of different models on two datasets, (a) the training loss of different models on Sport-1M varying with the number of iterations (b) the training loss of different models varying with the number of iterations on UCF-101 (see online version for colours)

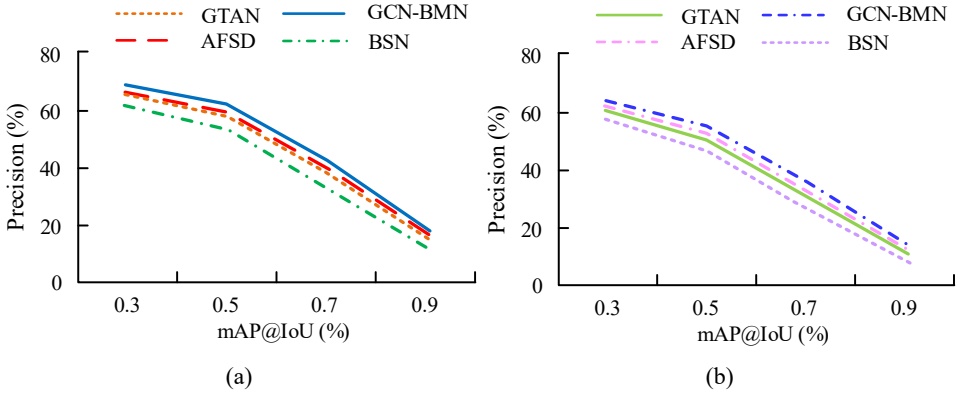


Figures 10(a) and 10(b) show the variation of training loss with iteration times for different models on two datasets. In the Sports-1M dataset, EF-GCN showed a 22.1%, 9.7%, and 5.7% improvement in training loss reduction with increasing iteration times compared to LSTM, ST-GCN, and DGNN models, respectively. In the UCF-101 dataset, EF-GCN gradually increased with the number of iterations z, and the degree of reduction in training loss compared to LSTM, ST-GCN, and DGNN models improved by 12.9%, 7.8%, and 4.8%, respectively. This indicated that the research model had better training effectiveness and convergence performance.

3.2 Verification of sports movement localisation method combining key points of the human body

In the above content, the sports action classification performance verification of the relevant models was completed. Next, the performance of GCN-BMN was experimentally verified. Due to the inconsistency of the objects targeted by the two models, a new dataset needed to be established for action localisation. The THUMOS dataset contained a large number of large-scale video datasets for action recognition and behaviour localisation, which were used to evaluate the performance of video action recognition and action detection algorithms. The VSRep dataset contained repetitive movements and key points of the human body, which could better reflect the real situation of sports movements. It is noteworthy that the datasets used for action classification (Sports-1M, UCF-101) and temporal localisation (VSRep, THUMOS) exhibit differences in action characteristics. Sports-1M and UCF-101 encompass a broad range of discrete, general sports actions, whereas VSRep focuses on highly repetitive, fine-grained movements. This domain discrepancy imposes stricter demands on the model's generalisation capability. We intentionally adopted this configuration to validate the ability of the EF-GCN and GCN-BMN models in learning universal spatio-temporal features. By measuring the performance differences under different IoU thresholds, the experiment accurately measured the accuracy and precision of the model's action localisation in the spatiotemporal dimension. Boundary sensitive network (BSN) is a model used in behaviour proposal networks to improve the accuracy of behaviour detection. Anchor free saliency-based detector (AFSD) is commonly-used for temporal action localisation tasks. Gaussian temporal awareness networks (GTAN) can accurately locate and classify actions in videos. The comparison results are shown in Figure 11.

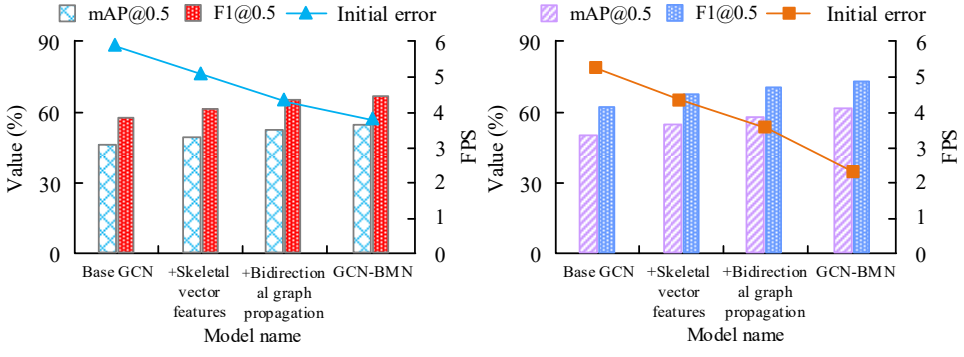
Figure 11 Comparison of the performance of various models in temporal action localisation, (a) comparison of temporal action localisation performance of different models in the VSRep dataset (b) comparison of timing action localisation performance of different models on the THUMOS dataset (see online version for colours)



Figures 11(a) and 11(b) show the comparison of temporal action localisation performance of various models on the VSRep and THUMOS datasets, respectively. Specific analysis showed that when the threshold was 0.9 on the VSRep dataset. The positioning accuracy of GCN-BMN improved by 51.2% compared to BSN, 16.5% compared to GTAN, and 9.5% compared to AFSD. On the THUMOS dataset, GCN-BMN improved by 67.1%

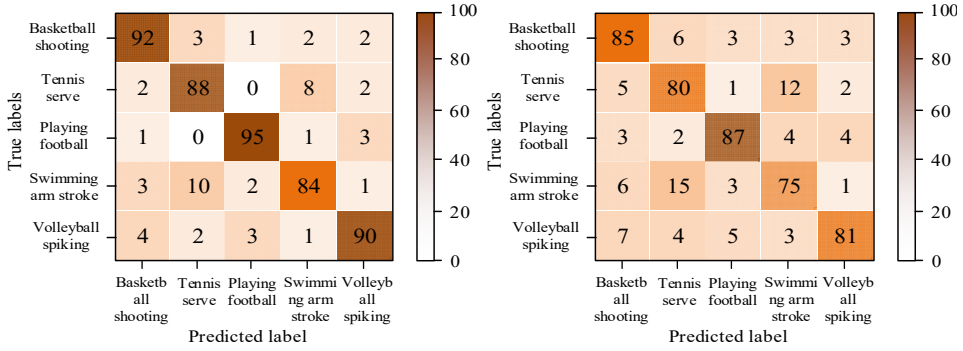
compared to BSN, 25.7% compared to GTAN, and 10.9% compared to AFSD. These data indicated that the research model had stronger accuracy and robustness in handling action localisation tasks in complex scenes. The ablation experiment was an effective method for evaluating the importance of each component or module in a model. By gradually removing or modifying certain components, the performance of the model was observed to determine the contribution of each part to the overall performance. The comparative study of ablation experiments is shown in Figure 12.

Figure 12 Ablation experiments on the VSRep and THUMOS datasets, (a) performance comparison of ablation experiments on the THUMOS dataset (b) ablation experiment analysis was conducted on the VSRep dataset (see online version for colours)



Figures 12(a) and 12(b) presented the ablation experiments of the model, from which it could be observed that performance gradually improved with the addition of modules. On the THUMOS dataset, compared to the initial GCN model, after incorporating skeletal vector features, the mAP and F1 scores increased by 3.6% and 3.3%, respectively, while the starting error frames decreased by 13.6%. With the further inclusion of bidirectional graph propagation, based on the previous improvements, the mAP and F1 scores increased by an additional 2.8% and 3.6%, respectively, and the starting error frames decreased by 27.1%. The final model achieved relative increases of 8.3% and 9.1% in mAP and F1 scores, respectively, with a reduction of 35.6% in starting error frames. On the VSRep dataset, after incorporating skeletal vector features, the mAP and F1 scores increased by 4.8% and 5.4%, respectively, while the starting error frames decreased by 17.3%. With the further inclusion of bidirectional graph propagation, based on the previous improvements, the mAP and F1 scores increased by an additional 8.3% and 8.3%, respectively, and the starting error frames decreased by 32.7%. The final model achieved relative increases of 11.8% and 11.0% in mAP and F1 scores, respectively, with a reduction of 55.8% in starting error frames. These results indicated that the research method had a significant effect in improving the accuracy and robustness of action localisation. To thoroughly analyse the performance differences among various models in action classification tasks, the study visually demonstrated the classification effects of each model through confusion matrices, as shown in Figure 13.

Figure 13 Comparison chart of confusion matrices between, (a) GCN-BMN model (b) BSN model (see online version for colours)



Figures 13(a) and 13(b) displayed the confusion matrices of the GCN-BMN and BSN models, respectively, for various sports action localisation and classification tasks. Through comparative analysis, it was found that, in the basketball shooting category, the accuracy of the research model improved by approximately 8.2% compared to the BSN model. In the tennis serve category, the improvement was 11.3%; in the soccer kicking category, it was 8.0%; in the swimming arm-stroke category, it was 9.3%; and in the volleyball spiking category, it was 9.9%. Overall, the GCN-BMN model demonstrated an average accuracy improvement of approximately 9.3% across all action categories compared to the BSN model, showcasing stronger localisation and classification accuracy as well as discriminative capability.

3.3 Verification of sports movement analysis method combining early fusion network structure and key points of human body

Finally, the performance of the proposed EFK-SAA model was studied and validated. The video inference for body pose and expression (VIBE) model combines Transformer and LSTM to simultaneously perform pose estimation and action recognition on videos. Therefore, the experiment compared various performance indicators and performance consumption of the above models to verify the practicality and accuracy of the model, as shown in Figure 14.

Figure 14(a) presented a comparison of the performance metrics between the EFK-SAA and VIBE models. The EFK-SAA model outperformed the VIBE model across all four metrics – accuracy, precision, recall, and F1-score – with improvements of 9.8%, 10.7%, 9.4%, and 10.1%, respectively. Figure 14(b) illustrated a comparison of resource utilisation between the EFK-SAA and VIBE models, revealing that the EFK-SAA model reduced GPU memory usage by 20.7% and system memory usage by 20.0%. In summary, the EFK-SAA model demonstrated superior performance metrics compared to the VIBE model while also being more resource-efficient. This indicated that the EFK-SAA model maintained high performance with lower resource requirements, rendering it more appropriate for environments with limited resources. The pursuit of high performance must be balanced with computational efficiency to achieve practical deployment. For an objective evaluation of this trade-off, the study compared

key hardware-independent metrics – parameter count, floating-point operations (FLOPs), and throughput – with the results summarised in Table 2.

Figure 14 Comparison of performance requirements and indicators between the EFK-SAA and VIBE models, (a) comparison of model performance indicators (b) comparison of model performance requirements (see online version for colours)

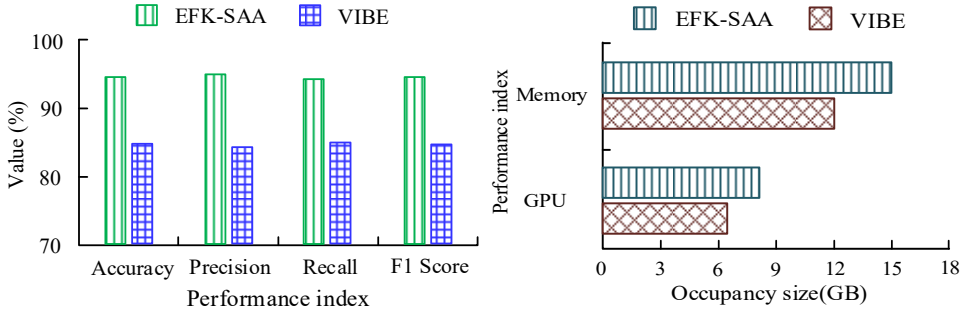


Table 2 Model efficiency and throughput comparison

Model	*Parameters (M)*	*FLOPs (G)*	*Throughput (FPS, BS = 8)*
VIBE	45.7	12.5	37.1
ST-GCN	32.1	9.2	51.3
DGNN	35.2	9.9	48.5

The data in Table 2 provides quantifiable evidence for the superior efficiency of the EFK-SAA model. When compared to existing graph convolutional networks, EFK-SAA demonstrates clear advantages. It uses 12% fewer parameters than ST-GCN and requires 15% fewer FLOPs. Furthermore, it achieves approximately 20% reduction in both parameters and FLOPs compared to the DGNN model. This indicates that the early fusion strategy is more effective in constructing compact and computationally efficient graph representations than the sequential or dual-stream approaches adopted by these established GCNs. This architectural efficiency directly translates to the highest inference throughput (69.4 FPS), surpassing ST-GCN by 35% and DGNN by 43%. These hardware-agnostic metrics confirm that the model's 'high efficiency' stems from its fundamental architectural innovations, making it particularly suitable for real-time applications. Finally, the experiment validated the analytical performance of the models for sports action analysis tasks using confusion matrices, as shown in Figure 15.

Figures 15(a) and 15(b) respectively show the confusion matrices of EFK-SAA and VIBE models in sports action analysis tasks. Through comparative analysis, it can be concluded that EFK-SAA improved the accuracy of running category analysis by 4.6% compared to VIBE, jumping category analysis by 10.0%, throwing category analysis by 3.66%, and hitting category analysis by 3.53%. This indicated that the EFK-SAA model had high accuracy and strong discriminative ability in action classification tasks. To evaluate the model's tolerance to upstream keypoint detection errors, the study simulated varying levels of detection inaccuracies by injecting Gaussian noise into the keypoint coordinates and observed the corresponding changes in model performance. The results are presented in Table 4.

Figure 15 Comparison of confusion matrices for different models Dels, (a) EFK-SAA model (b) VIBE model (see online version for colours)

Table 3 Model performance under different keypoint noise levels

Noise level (σ)	*Description*	*EFK-SAA mAP (%)*	*Δ mAP (percentage points)*
0	Clean (original)	76.5	-
0.02	Low noise	75.1	-1.4
0.05	Moderate noise	73.8	-2.7

According to Table 4, under low to moderate noise levels ($\sigma \leq 0.05$), the performance degradation of the EFK-SAA model is minimal (<3.5%). This indicates that the model does not overfit to ideal 'clean' data and demonstrates robust tolerance towards slight jitter or minor inaccuracies in keypoint coordinates. This noise-insensitive robustness is crucial for ensuring reliable performance in practical applications characterised by complex real-world conditions, such as occlusions and motion blur. To ensure the motion analysis model can adapt to motion capture data from different sources, its cross-device generalisation capability was validated. The experiment involved training the model on data from one type of device and conducting zero-shot testing on data from another device, thereby evaluating the model's robustness to systematic biases between devices. The performance comparison is shown in Table 4.

Table 4 Model performance under different keypoint noise levels

Training device	*Test device*	*ST-GCN*	*DGNN*	*EFK-SAA*
Vicon (high-end)	Kinect (consumer)	68.3	71.5	79.2
Kinect	Vicon	65.1	68.9	76.8
Average cross-device accuracy	-	66.7	70.2	78.0

The cross-device test results (Table 4) clearly demonstrate that the EFK-SAA model exhibits the strongest generalisation capability and robustness when handling keypoint data from different sources. Whether generalising from high-precision equipment to consumer-grade devices or vice versa, the model's performance degradation is significantly smaller than that of ST-GCN and DGNN models. This advantage primarily stems from the adopted early fusion strategy and data augmentation techniques like Neville interpolation, which guide the model to learn more universal spatiotemporal

feature representations rather than overfitting to device-specific noise or data characteristics. This validation confirms that the EFK-SAA model can better adapt to motion capture hardware of varying brands and specifications, substantially enhancing its deployment value and application potential in heterogeneous real-world environments.

4 Discussion

In recent times, with the ongoing advancement of deep learning technology, sports movement analysis has been widely studied as a key technology for improving athlete movement scores and preventing injuries. To address the issue of insufficient applicability in traditional analysis, a sports action analysis model combining early fusion network structure and human key points was proposed. The experiment outcomes indicated that on the Sports-1M dataset, the classification accuracy of the EF-GCN model was improved by 18.5%, 1.5%, and 3.4% respectively compared to the LSTM, ST-GCN, and DGNN models, and the accuracy was improved by 18.0%, 1.4%, and 3.3% respectively. In contrast, the general supervised machine learning classification method proposed by Worsey et al., although achieving lightweight deployment in athlete state monitoring, relied on a single sensor data and static feature extraction strategy, making it difficult to capture the spatiotemporal dynamic correlations and joint coordination patterns of sports movements (Worsey et al., 2021). By integrating human skeleton nodes and edge information through EF-GCN and introducing dynamic topology layers and keypoint weight distributions, research could more accurately capture human motion patterns, thereby achieving significant improvements in classification accuracy and precision, providing more reliable and efficient technical support for sports action classification. On VSRep, when the IoU threshold was 0.9, GCN-BMN improved localisation accuracy by 51.2%, 16.5%, and 9.5% compared to BSN, GTAN, and AFSD models, respectively. According to the ablation experiment on THUMOS, adding only skeletal vector features increased mAP by 3.6%, while further introducing bidirectional graph propagation increased mAP by 2.8%. The final mAP of the model relative to the initial GCN increased by 8.3%. While the study did not separately report the absolute number of false positive segments, the model's significantly higher precision provides direct and compelling evidence of its capability in false positive control. In action localisation tasks, precision is defined as the proportion of true positives among all positive samples predicted by the model (TP / (TP + FP)). Therefore, the substantial improvement in precision achieved by the model on the VSRep and THUMOS datasets compared to baseline models such as BSN and GTAN directly demonstrates a significant reduction in the number of erroneous proposals generated. The systematic review by Pu et al. pointed out that traditional trajectory analysis methods in football action localisation had boundary recognition errors of over 15% for emergency stop/direction change scenarios (Pu et al., 2024). By using a smooth frame discrimination module to adaptively partition action intervals, the localisation accuracy of football kicking categories was improved by 8.0%, which was significantly better than existing sports specific analysis models.

For the EFK-SAA model, when compared to the VIBE model, its performance improved by 9.8%, 10.7%, 9.4%, and 10.1% in terms of accuracy, precision, recall, and F1-score, respectively. Meanwhile, the GPU memory and system memory usage of the EFK-SAA model decreased by 20.7% and 20.0%, respectively. The VIBE model relied

on LSTM and transformer to capture long-term temporal dependencies, whereas the EFK-SAA model modelled structural changes in actions within the spatial dimension through a dynamic topology layer and bidirectional graph propagation, supplemented by temporal convolutions to extract local dynamics, thereby reducing the number of parameters. The sports action recognition method proposed by Nadeem et al. which was based on deep learning and clustering-based feature extraction algorithms, achieved certain results in action recognition. However, its generalisation capability was limited when dealing with action sequences of inconsistent lengths (Nadeem et al., 2021). This approach primarily depended on clustering algorithms for feature extraction and classification of actions, leading to a decline in accuracy and robustness when handling complex and variable sports action sequences. In contrast, the EFK-SAA method, through its early-fusion network architecture and Neville's interpolation method, was better able to adapt to action sequences of varying lengths, enhancing the model's generalisation capability and adaptability.

In summary, the EFK-SAA method proposed in the study achieved significant improvements in the classification accuracy and temporal localisation accuracy of sports movements compared to existing mainstream methods. This method not only performed well in core indicators, but also was more efficient in utilising computing resources. This provided a more powerful and practical analytical tool for refined sports training, technical movement assessment, and sports injury prevention.

5 Conclusions

In today's digital age, sports movement analysis, with the help of computer technology, provides powerful support for athletes' performance improvement, injury prevention, and rehabilitation. A new sports action analysis method combining early fusion network structure and human key points was proposed to address the problems of insufficient target differentiation in multiple scenarios, strong perspective dependence, and inconsistent action sequence length in existing technologies. By integrating human skeleton nodes and edge information, and introducing the Neville interpolation method and smooth frame discrimination module, the study aimed to improve the accuracy and stability of sports analysis. The results indicated that EFK-SAA had higher accuracy and efficiency in sports action classification and localisation tasks. However, the study still has the following limitations: Firstly, the model's performance relies on the accuracy of upstream pose estimation and remains relatively sensitive to keypoint errors in occluded scenarios. Secondly, the current method lacks biomechanical constraints, which may affect the plausibility of complex movement analysis. Additionally, the model is designed for fixed action categories and lacks the ability to adapt to emerging sports. Future research will focus on: developing representation learning methods more robust to pose estimation noise; integrating biomechanical principles into graph convolutional networks to enhance the physical plausibility of analysis; and exploring incremental learning frameworks to enable continuous model adaptation to new sports categories.

Funding

The research is supported by 2025 Guangdong Province Education Science Planning Project (Higher Education Special): Research on the Mechanism of Industry Education Integration Training for Sports Event Operations Talents from the Perspective of Artificial Intelligence (No. 2025GXJK0367).

Declarations

The authors have no competing interests to declare that are relevant to the content of this article.

References

Bao, L., Wang, Y., Song, X. and Sun, T. (2025) 'HGCGE: hyperbolic graph convolutional networks-based knowledge graph embedding for link prediction', *Knowl. Inf. Syst.*, Vol. 67, No. 1, pp.661–687, DOI: 10.1007/s10115-024-02247-8.

de Camargo, A.P. (2022) 'Backward and forward stability analysis of Neville's algorithm for interpolation and a pyramid algorithm for the computation of Lebesgue functions', *Numer. Algor.*, Vol. 89, No. 4, pp.1521–1531, DOI: 10.1007/s11075-021-01163-0.

Di Domizio, M. and Fabrizi, E. (2024) 'Ask me if I am happy: sport practice and life satisfaction in Italy', *Qual. Quant.*, Vol. 58, No. 6, pp.5865–5881, DOI: 10.1007/s11135-024-01921-x.

Dornaika, F. (2023) 'Deep, flexible data embedding with graph-based feature propagation for semi-supervised classification', *Cogn. Comput.*, Vol. 15, No. 1, pp.1–12, DOI: 10.1007/s12559-022-10056-w.

Endo, Y., Yan, X., Li, M., Akiyama, R., Brandl, C., Liu, J.Z., Hobara, R., Hasegawa, S., Wan, W., Novoselov, K.S. and Tang, W.X. (2023) 'Dynamic topological domain walls driven by lithium intercalation in graphene', *Nat. Nanotechnol.*, Vol. 18, No. 10, pp.1154–1161, DOI: 10.1038/s41565-023-01463-7.

Fan, T. and Lin, W. (2025) 'Deep learning-based thermal imaging analysis to diagnose abnormalities in sports buildings: smart cyber-physical monitoring sensors at the edge', *Tsinghua Sci. Technol.*, Vol. 30, No. 4, pp.1457–1473, DOI: 10.26599/TST.2023.9010130.

Geisen, M., Seifriz, F., Fasold, F., Slupczynski, M. and Klatt, S. (2024) 'A novel approach to sensor-based motion analysis for sports: piloting the Kabsch algorithm in volleyball and handball', *IEEE Sens. J.*, Vol. 24, No. 21, pp.35654–35663, DOI: 10.1109/JSEN.2024.3455173.

Hasanvand, M., Nooshyar, M., Moharamkhani, E. and Selyari, A. (2023) 'Machine learning methodology for identifying vehicles using image processing', *AIA*, April, Vol. 1, No. 3, pp.170–178, https://doi.org/10.47852/bonviewAIA3202833.

Li, S., Luo, H., Miao, Z., Wang, Z., Bao, Q. and Zhang, J. (2025) 'Foreground-driven contrastive learning for unsupervised human keypoint detection', *IEEE Photonics J.*, Vol. 17, No. 3, pp.1–14, DOI: 10.1109/JPHOT.2025.3567754.

Li, W., Lam, S., Wang, Y., Liu, C., Li, T., Kleesiek, J., Cheung, A.L.Y., Sun, Y., Lee, F.K-h., Au, K-h., Lee, V.H-f. and Cai, J. (2024) 'Model generalizability investigation for GFCE-MRI synthesis in NPC radiotherapy using multi-institutional patient-based data normalization', *IEEE J. Biomed. Health Inform.*, Vol. 28, No. 1, pp.100–109, DOI: 10.1109/JBHI.2023.3308529.

Li, X., Su, D., Chang, D., Liu, J., Wang, L., Tian, Z., Wang, S. and Sun, W. (2023) 'Multi-scale feature extraction and fusion net: research on UAVs image semantic segmentation technology', *J. ICT Standardization*, Vol. 11, No. 1, pp.97–116, DOI: 10.13052/jicts2245-800X.1115.

Luo, X., Jiang, R., Yang, B., Qin, H. and Hu, H. (2024) 'Air quality visualization analysis based on multivariate time series data feature extraction', *J. Visual.*, Vol. 27, No. 4, pp.567–584, DOI: 10.1007/s12650-024-00981-3.

Ma, H., Yang, Z. and Liu, H. (2022) 'Fine-grained unsupervised temporal action segmentation and distributed representation for skeleton-based human motion analysis', *IEEE Trans. Cybern.*, Vol. 52, No. 12, pp.13411–13424, DOI: 10.1109/TCYB.2021.3132016.

Ma, Q., Zhu, X., Zhao, X., Zhao, B., Fu, G. and Zhang, R. (2024) 'An equidistance index intuitionistic fuzzy C-means clustering algorithm based on local density and membership degree boundary', *Appl. Intell.*, Vol. 54, No. 4, pp.3205–3221, DOI: 10.1007/s10489-024-05297-1.

Malawski, F. (2021) 'Depth versus inertial sensors in real-time sports analysis: a case study on fencing', *IEEE Sens. J.*, Vol. 21, No. 4, pp.5133–5142, DOI: 10.1109/JSEN.2020.3036436.

Mishina, T., Okui, M. and Okano, F. (1995) 'A study on three-dimensional (spatio-temporal) noise characteristics of letterbox EPITV', *IEEE Trans. Broadcast.*, Vol. 41, No. 4, pp.113–120, DOI: 10.1109/TBC.1995.8952078.

Miyake, S. and Miyake, T. (2025) 'Force estimation of five fingers using infrared optical sensors and an IMU and its application to analysis of sports motion', *IEEE Sens. J.*, Vol. 25, No. 6, pp.10122–10133, DOI: 10.1109/JSEN.2025.3528649.

Movassagh, S., Fatehi, A., Sedigh, A.K. and Shariati, A. (2023) 'Kalman filter fusion with smoothing for a process with continuous-time integrated sensor', *IEEE Sens. J.*, Vol. 23, No. 7, pp.7279–7287, DOI: 10.1109/JSEN.2023.3244659.

Mumuni, A. and Mumuni, F. (2025) 'Data augmentation with automated machine learning: approaches and performance comparison with classical data augmentation methods', *Knowl. Inf. Syst.*, Vol. 67, No. 5, pp.4035–4085, DOI: 10.1007/s10115-025-02349-x.

Nadeem, A., Jalal, A. and Kim, K. (2021) 'Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy Markov model', *Multimed. Tools Appl.*, Vol. 80, No. 14, pp.21465–21498, DOI: 10.1007/s11042-021-10687-5.

Priyanka, S.S. and Kumar, T.K. (2023) 'Multi-channel speech enhancement using early and late fusion convolutional neural networks', *Signal Image Video Process.*, Vol. 17, No. 4, pp.973–979, DOI: 10.1007/s11760-022-02301-4.

Pu, Z.Z., Pan, Z.Q., Yi, H.W., Shi, J., Liu, B.Y., Chen, M., Ma, H. and Cui, Y.X. (2024) 'Football positioning and decision-making based on motion analysis and artificial intelligence: a systematic review', *IEEE/CAA J. Autom. Sin.*, Vol. 11, No. 1, pp.37–57, DOI: 10.1109/JAS.2023.123807.

Sasikaladevi, N., Pradeepa, S., Revathi, A., Vimal, S. and Crespo, R.G. (2024) 'Diagnosis of kidney cyst, tumor and stone from CT scan images using feature fusion hypergraph convolutional neural network (F2HCN2)', *Int. J. Mult. Comp. Eng.*, Vol. 22, No. 5, pp.35–46, DOI: 10.1615/IntJMultCompEng.2023048245.

Shakrani, K.V., Kanyangarara, N.M., Parowa, P.T., Gupta, V. and Kumar, R. (2022) 'A deep learning model for face recognition in presence of mask', *Acta Informatica Malaysia*, Vol. 6, No. 2, pp.43–46, https://doi.org/10.26480/aim.02.2022.43.46.

Su, S., Gao, T., Zhu, Y., Fang, X. and Fan, T. (2025) 'Rockburst prediction via multiscale graph convolutional neural network', *Rock Mech. Rock Eng.*, Vol. 58, No. 1, pp.659–677, DOI: 10.1007/s00603-024-04182-0.

Worsey, M.T.O., Espinosa, H.G., Shepherd, J.B. and Thiel, D.V. (2021) 'One-size-fits-all: supervised machine learning classification in athlete monitoring', *IEEE Sens. Lett.*, Vol. 5, No. 3, pp.1–4, DOI: 10.1109/LSENS.2021.3060376.

Xiao, X., Xu, W., Tang, Y., Li, W., Dong, D., Shangguan, Y. and Huang, S. (2023) 'Improved loss minimization control based on time-harmonic equivalent circuit for linear induction motors adopted to linear metro', *IEEE Trans. Veh. Technol.*, Vol. 72, No. 7, pp.8601–8612, DOI: 10.1109/TVT.2023.3244602.

Xue, L., Lu, N., Chen, C., Hu, T. and Jiang, B. (2023) 'Attention mechanism based multi-scale feature extraction of bearing fault diagnosis', *J. Syst. Eng. Electron.*, Vol. 34, No. 5, pp.1359–1367, DOI: 10.23919/JSEE.2023.000129.

Yu, Q. and Bo, L. (2024) 'Influence of body movements of sports model on the advertising effect of sports clothing print network in the internet context', *Curr. Psychol.*, September, Vol. 43, No. 36, pp.28738–28755, DOI: 10.1007/s12144-024-06501-8.

Yu, W., Li, Y., Fan, C., Fu, D., Zhang, C., Chen, Y., Qian, M., Liu, J. and Liu, G. (2024) 'Precipitation nowcasting leveraging spatial correlation feature extraction and deep spatio-temporal fusion network', *Earth Sci. Inform.*, Vol. 17, No. 5, pp.4739–4755, DOI: 10.1007/s12145-024-01412-5.

Zamani, S.A. and Baleghi, Y. (2023) 'Early/late fusion structures with optimized feature selection for weed detection using visible and thermal images of paddy fields', *Precis. Agric.*, Vol. 24, No. 2, pp.482–510, DOI: 10.1007/s11119-022-09954-8.

Zhang, J., Chen, Z. and Tao, D. (2021) 'Towards high performance human keypoint detection', *Int. J. Comput. Vis.*, Vol. 129, No. 9, pp.2639–2662, DOI: 10.1007/s11263-021-01482-8.

Zhang, Y., Huang, D., Lv, G. and Zhao, H. (2024) 'A new real-time positioning correction system based on an inherited sampling algorithm with an FPGA accelerator', *IEEE IoT J.*, Vol. 11, No. 8, pp.14914–14923, DOI: 10.1109/JIOT.2023.3345664.