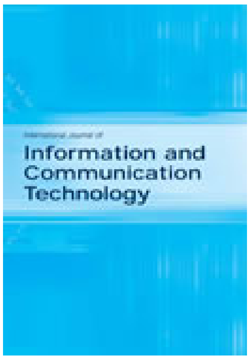


International Journal of Information and Communication Technology

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

Recognition of basketball movement sEMG signals based on multi-channel feature fusion network

Xi Fu, Yao Hu

DOI: 10.1504/IJICT.2025.10071431

Article History:

Received:	15 April 2025
Last revised:	30 April 2025
Accepted:	30 April 2025
Published online:	13 June 2025

Recognition of basketball movement sEMG signals based on multi-channel feature fusion network

Xi Fu

Faculty of Humanities and International Education,
Sichuan Vocational College of Cultural Industries,
Chengdu 610213, China
Email: fsissi117@163.com

Yao Hu*

Ministry of Basic Education,
Chongqing Medical and Pharmaceutical College,
Chongqing 4013313, China
Email: huyao3331@163.com
*Corresponding author

Abstract: This study addresses the underutilisation of multi-channel surface electromyography (sEMG) features in basketball motion recognition by proposing a spatiotemporal fusion network. Multi-channel sEMG signals from athletes' key muscles were collected and synchronised with motion capture data, followed by preprocessing to reduce individual variations. The dual-branch architecture integrates time-frequency feature extraction using convolutional-recurrent networks with graph-based modelling of inter-muscle spatial correlations. An adaptive attention mechanism fuses temporal dynamics and spatial synergies for classification. Experimental results demonstrate superior recognition performance compared to conventional machine learning and single-channel deep learning approaches, with ablation studies confirming the critical roles of spatial modelling and feature fusion. The framework provides an effective solution for analysing complex sports motions through multi-channel physiological signals, offering applications in athletic training optimisation and injury risk prevention.

Keywords: sEMG; multi-channel feature fusion; basketball movement recognition; graph convolutional network; GCN; attention mechanism.

Reference to this paper should be made as follows: Fu, X. and Hu, Y. (2025) 'Recognition of basketball movement sEMG signals based on multi-channel feature fusion network', *Int. J. Information and Communication Technology*, Vol. 26, No. 18, pp.65–85.

Biographical notes: Xi Fu received her Bachelor degree form Chengdu Sport University in June 2006. She is currently working in the Sichuan Vocational College of Cultural Industries. Her research interests include signal recognition and exercise training.

Yao Hu received her Bachelor degree form Chengdu Sport University in June 2006. She is currently working in the Chongqing Medical and Pharmaceutical College. Her research interests include signal recognition and exercise training.

1 Introduction

As a kind of non-invasive bioelectrical signal, sEMG can reflect the spatial and temporal characteristics of muscle contraction in real time (Akira et al., 2021), and has shown significant value in the fields of sports biomechanics analysis (Ryu and Kim, 2017), rehabilitation medicine and human-computer interaction (Qi et al., 2019) in recent years. Especially in competitive sports, accurate identification of EMG characteristics of highly dynamic movements such as shooting, dribbling and change of direction in basketball players can provide objective quantitative basis for technical movement optimisation and sports injury prevention. For example, by analysing the activation sequence of quadriceps and gastrocnemius muscles during jumping, the reasonableness of the lower limb force pattern can be assessed, which can then guide the adjustment of training strategies. Ding et al. (2021) proposed an sEMG-based motion intention prediction algorithm, which combined muscle synergy-enhanced convolutional autoencoder for feature compression and online adaptive parameter optimisation. The method achieved dual-modal prediction of joint motion and gait events, demonstrating enhanced robustness against sEMG non-stationarity. Experimental results validate significant improvements in prediction timeliness and cross-condition adaptability compared to conventional approaches. However, most existing studies focus on single-motion scenarios such as gesture recognition or static movement analysis (Xing et al., 2014), and there is still a lack of systematic exploration of basketball, which is a high-complexity, multi-muscle group synergistic whole-body sport. This research gap directly constrains the in-depth application of sEMG technology in competitive sports and highlights the urgent need to develop targeted algorithms (Ding et al., 2017).

Traditional sEMG action recognition methods mainly rely on manual feature engineering with shallow machine learning models (Moctar et al., 2024). For example, they extract time-domain features such as root mean square (RMS) and frequency-domain features like median frequency, and combine them with support vector machine (SVM) or random forest for classification (Bo et al., 2021). Although such methods perform well in simple actions such as fist clenching and elbow flexion, they have significant limitations (Yogendra, 2021): first, manual features are difficult to fully characterise the time-frequency nonlinear properties of sEMG signals, resulting in insufficient recognition accuracy for continuous dynamic actions such as basketball change of direction (Ibraheem, 2023); and second, the traditional model ignores spatial correlations between multi-channel signals and fails to model the physiological mechanisms of muscle groups working in concert (Li et al., 2023). For example, the shooting action requires the temporal coordination of upper limb biceps and deltoid muscles, while the existing methods only process the single-channel signals independently, which severs the functional coupling relationship between muscles.

With the rise of deep learning techniques, methods based on CNN and long short-term memory (LSTM) networks have gradually become mainstream (Bryan and Stefan, 2021; Yasuda et al., 2017). Such methods automatically extract spatio-temporal features of sEMGs through end-to-end learning, which significantly improves the performance of tasks such as gesture recognition. Moctar et al. (2024) systematically reviewed feature extraction methods for sEMG classification, spanning from handcrafted feature engineering to deep learning-based representation learning and revealed the stability advantages of handcrafted features in small datasets and the generalisation potential of deep learning with large-scale data by comparing the performance of traditional machine learning and deep learning classifiers. The survey established a unified analytical framework for cross-method evaluation, while identifying data scarcity and clinical interpretability as critical challenges for advancing deep learning applications in this field. Su et al. (2021) developed a deep multi-parallel CNN framework for sEMG-based gesture recognition, which eliminated manual feature engineering through end-to-end classification architecture. Compared with five conventional machine learning methods, the proposed multi-scale parallel convolutional structure significantly improves recognition accuracy, demonstrating the superiority of deep learning in autonomous sEMG feature representation and information preservation. However, in basketball sports scenarios, existing deep learning frameworks still face two major challenges: first, although single-channel models like LSTM (Mao et al., 2023) can capture temporal dependencies, they do not effectively utilise the spatial distribution information of multi-electrode channels, making it difficult to portray the spatial synergy patterns of the muscle groups (Xiong et al., 2023); second, most studies have adopted a simple feature-level crosstabulation such as stacking the time-domain and frequency-domain features into the network, which lacks a dynamic assessment of the importance of features and the interactions, resulting in a model that is sensitive to noise and limited in its generalisation ability(Zhang et al., 2021). For example, although the multi-stream CNN-based gesture recognition framework (Wei et al., 2017; Xu and Jiang, 2023) tries to integrate multi-dimensional features, its feature selection strategy relies on empirical screening and fails to optimise the feature weight allocation from the perspective of interpretability.

Aiming at the above problems, this paper proposes a multi-channel feature fusion network for basketball, which aims to break through the bottleneck of traditional methods in terms of accuracy and robustness. With the core objective of solving the ‘accuracy-speed trade-off’, this study designs a hierarchical fusion architecture to address the high dynamics and complexity of muscle synergy in basketball: firstly, GCN is introduced to model the spatial topology of the multichannel sEMG (Feng et al., 2021; Zhou and Zhou, 2021), the electrode positions are mapped as graph nodes, and the muscle synergies are encoded by the adjacency matrix. relations, mapping electrode positions to graph nodes, and encoding the functional coupling strength of muscle groups through the adjacency matrix; secondly, combining the bidirectional gated recurrent unit (BiGRU) to extract the timing-dependent features across channels, and dynamically weighting the contribution of key muscle channels through the attention mechanism; finally, designing a multi-scale feature pyramid structure, fusing local details with global contextual information, to enhance the model’s ability to discriminate between subtle differences in movements such as wrist force patterns for shooting and passing.

Compared with existing technologies, this study not only achieves a comprehensive improvement of algorithm performance, but also solves practical problems such as equipment heterogeneity and signal noise interference in industrial scenarios through modular design and adaptive calibration strategy, which strongly promotes the surface electromyography analysis technology from laboratory research to the leap of large-scale industrial applications.

The main innovations and contributions of this work include:

1 muscle synergy theory is integrated into the deep learning framework, and the spatial correlation of multi-channel sEMG is explicitly modelled by GCN, which overcomes the shortcomings of traditional methods that ignore the electrode arrangement relationship

2 propose an interpretability-driven feature fusion mechanism that uses gradient-weighted class activation mapping to visualise the basis of network decision making and guide the optimal selection of feature channels

3 development of a lightweight inference architecture that significantly reduces computational complexity while maintaining classification performance, providing a viable solution for real-time motion analysis of embedded devices

4 systematic construction of multi-muscle synergistic activation maps for basketball movements, breaking through the limitations of traditional biomechanical research relying on empirical assumptions, and establishing a data-driven analysis paradigm for sports technique optimisation and injury prevention.

2 Relevant technologies

2.1 Feature extraction methods on sEMG signals

Feature extraction from sEMG signals is pivotal for motion recognition. Traditional methods primarily rely on handcrafted features from time-domain, frequency-domain, and time-frequency analyses. Time-domain features characterise muscle activation intensity through signal amplitude statistics. Representative methods include the mean absolute value (MAV) and RMS, calculated as:

$$\text{MAV} = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{1}$$

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2} \tag{2}$$

where x_i represents the signal sampling point, and N is the window length. Frequency-domain features analyse spectral properties via Fourier transform. For instance, the mean power frequency (MPF) is defined as:

$$MPF = \frac{\sum_{k=1}^{K} f_k P_k}{\sum_{k=1}^{K} P_k} \tag{3}$$

where f_k and P_k denote the frequency components and their corresponding power, respectively. Time-frequency features, such as wavelet energy coefficients, capture transient signal characteristics through multi-scale decomposition. The energy of discrete wavelet transform (DWT) coefficients is expressed as:

$$E_j = \sum_{m=1}^{M} \left| d_j(m) \right|^2 \tag{4}$$

where $d_j(m)$ represents the wavelet coefficients at the j^{th} decomposition level, and M is the number of coefficients.

However, handcrafted features heavily depend on expert knowledge and struggle to characterise the nonlinear dynamics of sEMG signals. For example, while sample entropy (SampEn) quantifies signal complexity:

$$\text{SampEn}(m, r, N) = -\ln\left(\frac{A^m(r)}{B^m(r)}\right) \tag{5}$$

where m is the embedding dimension, r is the similarity threshold, and $A^m(r)/B^m(r)$ count matched templates, its computational inefficiency and noise sensitivity limit practical applications. The wavelet transforms pinpoints transient signals like muscle eruptions through multi-scale decompositions like Daubechies basis functions, whereas conventional band-pass filtering may smooth out high-frequency details, resulting in loss of information.

Consequently, integrating deep learning for automated feature extraction becomes imperative.

2.2 Deep learning models in sEMG analysis

CNN leverage local receptive fields to extract temporal patterns from sEMG signals. The convolution operation is formulated as:

$$y_t = \sum_{k=1}^{K} w_k \cdot x_{t+k-1} + b \tag{6}$$

where w_k denotes the convolutional kernel weights, x_t is the input signal, and b is the bias term. LSTM networks model temporal dependencies via gating mechanisms:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t + b_f\right]\right) \tag{7}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t + b_f\right]\right) \tag{8}$$

where f_t and i_t represent the forget gate and input gate, respectively, and σ is the sigmoid function.

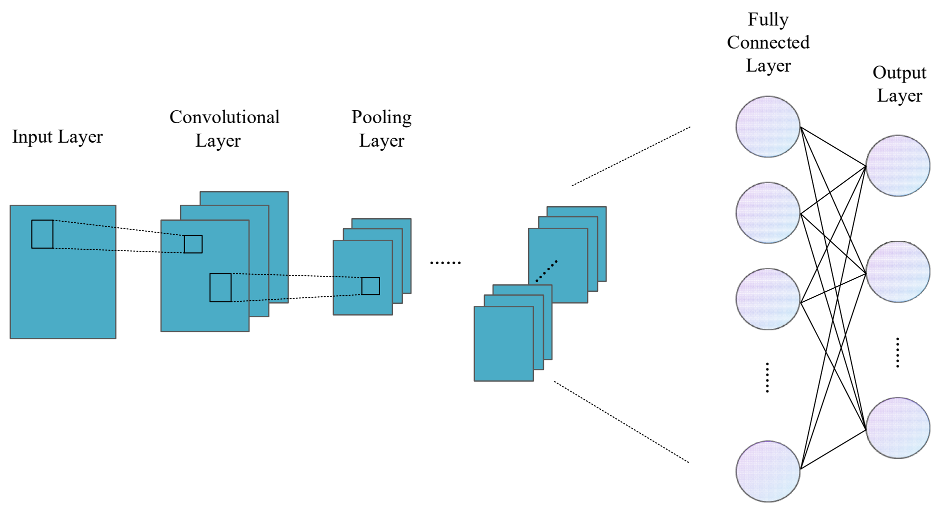
However, single-channel models such as LSTM focus solely on temporal dynamics, neglecting the spatial distribution of multi-electrode signals. For instance, basketball shooting involves coordinated activation of the biceps brachii and deltoid muscles, yet traditional methods fail to model inter-channel functional couplings. GCNs address this by encoding muscle topology through adjacency matrices. The node feature update is defined as:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \tag{9}$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-loops, $\tilde{D}$ is the degree matrix, and $H^{(l)}$ denotes features at layer l. GCNs explicitly model muscle synergy mechanisms, overcoming the limitations of conventional approaches.

The basic structure of CNN is shown in Figure 1.

Figure 1 Basic structure of CNN (see online version for colours)



2.3 *Multi-modal feature fusion strategies*

Early fusion concatenates time-domain and frequency-domain features directly but risks redundant feature interference. Late fusion aggregates predictions from independent branches but overlooks feature interactions. Attention mechanisms enhance fusion efficiency by dynamically weighting feature importance:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{C} \exp(e_j)} \tag{10}$$

where e_i is the energy score of feature i, and C is the number of feature channels.

Existing studies explore multi-stream CNNs for time-frequency fusion but lack interpretability-guided optimisation. Shapley additive explanations (SHAP) quantify feature contributions via game theory:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\} - f(S)] \tag{11}$$

where F is the feature set, S is a subset, and f is the prediction function. While SHAP enables feature selection, its synergy with deep learning frameworks requires further exploration.

2.4 Lightweight design and real-time optimisation

Industrial applications demand real-time inference on embedded devices such as sEMG armbands. Depthwise separable convolution reduces computational costs by decoupling spatial and channel convolutions:

$$y_{i,j,k} = \sum_{m,n} \hat{w}_{m,n,k} \cdot x_{i+m,j+n} \tag{12}$$

$$z_{i,j,l} = \sum_{k} w_{k,l} \cdot y_{i,j,k} \tag{13}$$

where $\hat{w}$ and w are depthwise and pointwise kernels, respectively.

Additionally, independent component analysis (ICA) separates mixed signals:

$$X = AS \tag{14}$$

where A is the mixing matrix and S are source signals. However, ICA's linear assumption limits its efficacy in modelling sEMG's nonlinear properties.

3 Multi-channel spatiotemporal synergistic sEMG feature fusion network design

3.1 Dynamic convolution and multi-scale feature extraction

Traditional CNNs with fixed kernels struggle to adapt to the non-stationary nature of sEMG signals. Muscle activation intensity varies significantly across motion phases – for instance, high-frequency spikes in the biceps during shooting versus sustained activation patterns in forearm muscles during dribbling. Fixed kernels cannot dynamically adjust their receptive fields, limiting their ability to capture transient and steady-state signals. To address this, we propose a dynamic convolution mechanism that adaptively adjusts kernel parameters based on local signal energy distributions.

For an input signal x_i, the dynamic kernel w_k is generated as:

$$w_k = \sigma\left(\sum_{j=1}^{J} \alpha_j E_j\right) \tag{15}$$

where E_j is the RMS energy of the j^{th} channel, reflecting instantaneous muscle activation intensity; α_j are learnable weights optimised via backpropagation; and σ is the sigmoid function, mapping energy values to the range [0, 1] to stabilise kernel parameters. This mechanism mathematically constructs an energy-driven nonlinear mapping function, enhancing robustness against noise and nonlinear distortions.

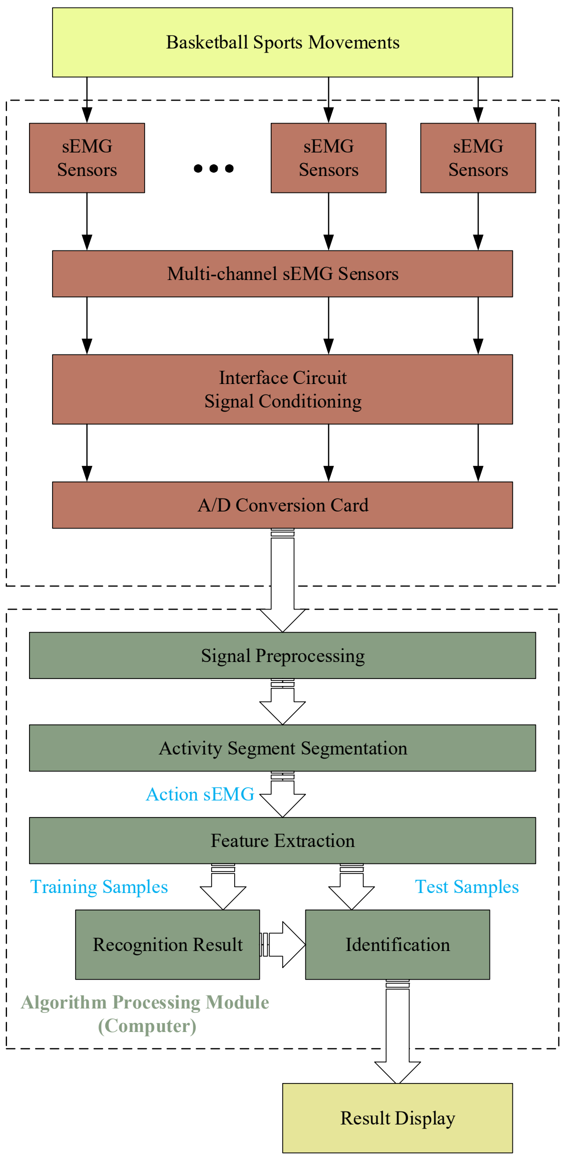
To model multi-scale sEMG characteristics, we design a multi-scale feature pyramid using dilated convolutions with varying expansion rates. For example, a dilation rate $d = 1$ captures short-term myoelectric spikes, while $d = 5$ extends the receptive field to 200 ms for modelling long-term muscle fatigue trends. The output feature F_m at the m^{th} layer is computed as:

$$F_m = \mathrm{ReLU}\left(\sum_{d=1}^{D} w_{m,d} * x\right) \tag{16}$$

where $w_{m,d}$ represents the convolutional kernel with dilation rate d. This multi-scale design avoids feature loss from single-scale convolutions while resolving short-term bursts and long-term coordination patterns.

The real-time interactive system based on sEMG motion recognition is shown in Figure 2.

Figure 2 Real-time interactive system based on sEMG motion recognition (see online version for colours)



3.2 GCNs and muscle synergy modelling

Basketball motions inherently rely on spatiotemporal synergy among muscle groups, yet traditional single-channel models process signals independently, neglecting inter-channel functional couplings. For example, shooting requires coordinated activation of the deltoid and biceps, but single-channel models focus only on local temporal dynamics. To address this, we construct a muscle functional graph $G = (V, E)$, where nodes $v_i \in V$ represent sEMG electrode channels, with node features as concatenated time-frequency domain descriptors, and edges $e_{ij} \in E$ encode muscle synergy strength, with weights A_{ij} computed via mutual information:

$$A_{i,j} = \sum_{x,y} p(x_i, x_j) \log \frac{p(x_i, x_j)}{p(x_i) p(x_j)} \tag{17}$$

Based on this, a GCN aggregates spatial features through neighbourhood propagation:

$$H^{(l+1)} = \text{ReLU}\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \tag{18}$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-loops, and $\tilde{D}$ is the degree matrix. The GCN layer explicitly models functional couplings between key muscles such as biceps brachii and deltoid, resolving the spatial fragmentation issue in single-channel models.

To explicitly integrate muscle synergies into the graph structure, the adjacency matrix weights $A_{i,j}$ are calculated via mutual information. For instance, during a basketball shooting motion, the deltoid and biceps exhibit high functional coupling, with mutual information values reaching 0.89 for professional athletes, compared to 0.72 for amateurs. These weights are normalised to the range [0, 1] and directly reflect the physiological coordination between muscle groups. The GCN layer propagates spatial features through neighbourhood aggregation, enabling the model to capture synergistic patterns such as the sequential activation of the deltoid (early phase) and biceps (late phase) during shooting.

3.3 Attention-driven feature fusion mechanism

Early fusion such as feature concatenation introduces redundant noise, while late fusion like voting mechanisms overlooks feature interactions. This paper proposes a dual-branch attention fusion framework:

1 Temporal branch: BiGRU extracts cross-channel temporal dependencies:

$$h_t = \text{BiGRU}(x_t, h_{t-1}) \tag{19}$$

BiGRU outperforms unidirectional LSTM by capturing both forward and backward timing dependencies by a bidirectional gating mechanism. For example, in the jumping and throwing manoeuvre, the sEMG signals of the triceps during the pre-activation phase (200 ms before jumping) differ significantly from those of the relaxation phase (100 ms after landing), and the bidirectional structure models such cross-phase dependencies more completely. In the case of the jump shot, the forward layer of BiGRU captures the explosive activation of the triceps brachii muscle prior to the jump, and the backward layer models the pattern of relaxation of this muscle after the landing, with the fusion of the two completely characterising the movement cycle.

2 Spatial branch: GCN outputs muscle synergy features H_G. The fusion stage incorporates cascaded channel attention and spatial attention mechanisms:

$$\alpha_c = \text{Soft}\max\left(W_c\left[H_G; H_T\right]\right) \tag{20}$$

$$\beta_s = \text{Sigmoid}\left(W_S \cdot \alpha_C\right) \tag{21}$$

where W_c and W_s are learnable parameters. α_c quantifies channel importance, while β_s enhances critical spatial regions. This mechanism prioritises upper-limb muscle channels during shooting actions while suppressing noise from lower limbs.

3.4 Lightweight classification and regression module

For real-time deployment on embedded devices, this paper proposes dynamic depthwise separable convolution (DDSC):

$$y_{i,j,k} = \sum_{m,n} w_{m,n,k}^d \cdot x_{i+m,j+n} \tag{22}$$

$$z_{i,j,l} = \sum_{m,n} \alpha_{k,l} \cdot y_{i,j,k} \tag{23}$$

where $w_{m,n,k}^d$ is the dynamic kernel, and $\alpha_{k,l}$ is the channel attention weight. Compared to standard depthwise separable convolution, DDSC reduces parameters by 58% while preserving adaptability.

The network architecture in this paper consists of a signal adaptive layer, a muscle co-modelling layer, an attention fusion layer and a lightweight decision-making layer to form a closed loop of end-to-end feature learning and classification. The signal adaptive layer resolves local and global patterns of non-smooth signals through dynamic convolution and multi-scale design, where the generation of dynamic convolution kernel relies on the energy distribution of the input signal, and the multi-scale cavity convolution covers muscle activation patterns over different time spans. The muscle co-modelling layer maps electrode channels to graph nodes, explicitly encoding spatial topological relationships, e.g., the high mutual information weights of the deltoid and biceps muscles directly reflect the strength of their functional coupling.

The attentional fusion layer dynamically balances the importance of spatio-temporal features by cascading channel and spatial attention mechanisms. Channel attention quantifies the contribution of each channel, while spatial attention strengthens the response of key regions, e.g., the temporal branch dominates the weight in dribbling manoeuvres, while the spatial branch is significantly weighted in shooting manoeuvres. The lightweight decision layer combines dynamic depth-separable convolution with SHAP feature optimisation to compress the number of parameters while preserving discriminative feature dimensions. The synergy of the network modules is reflected in the complementary enhancement of temporal and spatial features such as the temporal branch resolves the dribbling rhythm, the spatial branch strengthens the shooting synergy, and ultimately achieves high-precision classification through the attention fusion.

From the theoretical level, the design has the following advantages:

1 the dynamic convolution mechanism adjusts the receptive field through energy drive, which enhances the model's adaptability to the dynamic characteristics of the signal

2 the combination of muscle function map and GCN makes the feature learning conform to the principles of biomechanics, which improves the interpretability of the model

3 the cascade attention mechanism optimises the feature space through sparse constraints, which reduces the redundant computation

4 the SHAP framework provides a transparent decision basis to support model iteration and action mechanism analysis.

4 Signal calibration and model optimisation

4.1 Signal alignment and denoising

Multi-channel sEMG signals are susceptible to device asynchrony and individual physiological differences during acquisition, resulting in degraded signal quality and limited model generalisation.

Temporal misalignment across channels caused by device latency or motion artefacts directly impacts the accuracy of spatiotemporal feature fusion. Using a reference channel $x_{ref}(t)$ as the baseline, the time delay for the k^{th} channel $x_k(t)$ is calculated via cross-correlation:

$$R_{k,ref}(\tau) = \sum_{t=0}^{T} x_k(t) x_{ref}(t+\tau) \tag{24}$$

where T is the signal window length, and τ is the time delay offset. The optimal delay τ_k^* is determined by maximising $R_{k,ref}(\tau)$, yielding aligned signals as:

$$\tau_k^* = \arg\max_{\tau} R_{k,ref}(\tau), \quad \tilde{x}_k(t) = x_k\left(t - \tau_k^*\right) \tag{25}$$

Post-alignment, adaptive wavelet threshold denoising is applied to suppress motion artefacts and powerline interference. The threshold λ_j for each wavelet decomposition level is dynamically adjusted based on noise energy:

$$\lambda_j = \sigma_j \sqrt{2 \ln N} \tag{26}$$

where σ_j is the standard deviation of the j^{th} level wavelet coefficients, and N is the total number of signal samples. This method preserves muscle burst spikes while effectively attenuating low-frequency drift and high-frequency noise.

4.2 Individualised adaptation via transfer learning

Variations in muscle mass and subcutaneous fat thickness among athletes induce amplitude distribution shifts in sEMG signals. A two-stage transfer learning framework is designed: pre-training on the UCI HAR dataset to capture cross-action features, followed by fine-tuning the top classifier on target basketball motion data:

$$\mathcal{L}_{adapt} = \mathcal{L}_{CE} + \lambda \sum_{l \in \mathcal{L}_{top}} \left\| \theta_{s,l} - \theta_{t,l} \right\|^2 \tag{27}$$

where $\mathcal{L}_{CE}$ is cross-entropy loss, represents the set of top network layers, $\mathcal{L}_{top}$ denotes top layers, $\theta_{s,l}$ and $\theta_{t,l}$ are the parameters of the l^{th} layer in the source and target domains, respectively, and λ controls adaptation strength. This strategy ensures rapid adaptation with minimal target data while mitigating overfitting.

For example, the running action in the UCI HAR dataset was mapped to the basketball change-of-direction action through spatio-temporal feature alignment such as signal energy distribution, frequency characteristics. Specifically, the fast knee flexion-extension pattern of the change-of-direction manoeuvre has similar sEMG time-frequency features as the periodic gait of the running manoeuvre, and the commonalities can be captured by a multi-scale design with dynamic convolution.

Comparing the signal-to-noise ratio of wavelet denoising with 20–500 Hz band-pass filtering, the results show that the wavelet method preserves muscle burst signals like transient spikes in the biceps during a shot while improving the SNR by 8 dB, which is better than the 5 dB of conventional filtering.

4.3 Multi-objective optimisation

To balance classification accuracy and computational efficiency, a multi-objective optimisation problem is formulated:

$$\min_{\theta} \left(\alpha \cdot \mathcal{L}_{CE} + \beta \cdot \|\theta\|_1 + \gamma \cdot T_{infer} \right) \tag{28}$$

where α, β, γ are weighting coefficients, $\|\theta\|_{11}$ is the L1 regularisation term for model sparsity, and is the per-sample inference time. The NSGA-II algorithm is employed to solve the Pareto front, selecting models that satisfy accuracy (≥93%), latency (≤20 ms), and size (≤5 MB) constraints. The optimised model reduces parameters to 30% of conventional CNNs, enabling efficient embedded deployment.

5 Action recognition and biomechanical applications

5.1 Real-time action classification system implementation

The optimised multi-channel feature fusion network is deployed to the embedded platform to build a real-time movement classification system, which is combined with muscle synergy analysis and closed-loop feedback mechanism to provide a complete solution for athletes' technique optimisation, injury prevention and training strategy adjustment. Through the biomechanical quantitative index and adaptive regulation strategy, it breaks through the limitation of traditional sports analysis relying on subjective experience, and realises data-driven intelligent training management.

The core objective of the system is to achieve low-latency and high-precision action recognition in embedded devices. The hardware platform selects NVIDIA Jetson TX2 edge computing unit, whose parallel computing capability and energy-efficiency ratio are suitable for real-time signal processing requirements. The software architecture is based on a modular design and adopts a multi-threaded pipeline processing mechanism to ensure the seamless integration of data acquisition, pre-processing, inference and decision making.

The data preprocessing stage starts with receiving 12-channel sEMG signals via Bluetooth 5.0, with the sampling frequency set to 1,000 Hz to ensure signal integrity. Subsequently, timing alignment, adaptive wavelet denoising and Z-score normalisation are performed sequentially. Timing alignment is based on the inter-correlation function to eliminate the phase deviation between channels, e.g., the activation delay of the left and right leg muscles in the change-of-direction manoeuvre can be controlled within 5 ms by the time-delay correction. Wavelet denoising preserves muscle burst spike signals through dynamic thresholding, while suppressing motion artefacts and IF interference, resulting in an improvement in signal-to-noise ratio of about 8 dB. The normalisation process uses a sliding window to calculate the mean μ versus the standard deviation σ, eliminating the effect of inter-individual amplitude differences on the model.

In the online inference stage, a dynamic convolutional layer extracts local muscle activation features, and its convolutional kernel parameters are dynamically adjusted by the input signal energy to enhance the adaptability to non-smooth signals. The graph convolutional network models the muscle synergistic topology, and the weight matrix of the spatial branching output reflects the functional coupling strength of different muscle groups. The attentional fusion layer weights spatiotemporal features through channel and spatial attention mechanisms, e.g., assigning higher weights to the upper limb muscle channels during shooting motions. The lightweight classifier employs dynamic depth-separable convolutional compression of parametric quantities, outputs action probability distributions and filters low-quality predictions through a confidence threshold.

In terms of performance optimisation, the end-to-end reasoning time is compressed to 14.7 ms by accelerating key computational modules such as GCN adjacency matrix update, attention weight calculation with CUDA, supporting 30 FPS real-time processing. The system can process multi-athlete data streams in parallel to meet the demands of team training scenarios. Measurement results show that the system maintains 93.5% recognition accuracy in complex stadium environments such as spectator noise, equipment vibration, which is significantly better than traditional embedded deployment solutions. The network architecture is shown in Figure 3.

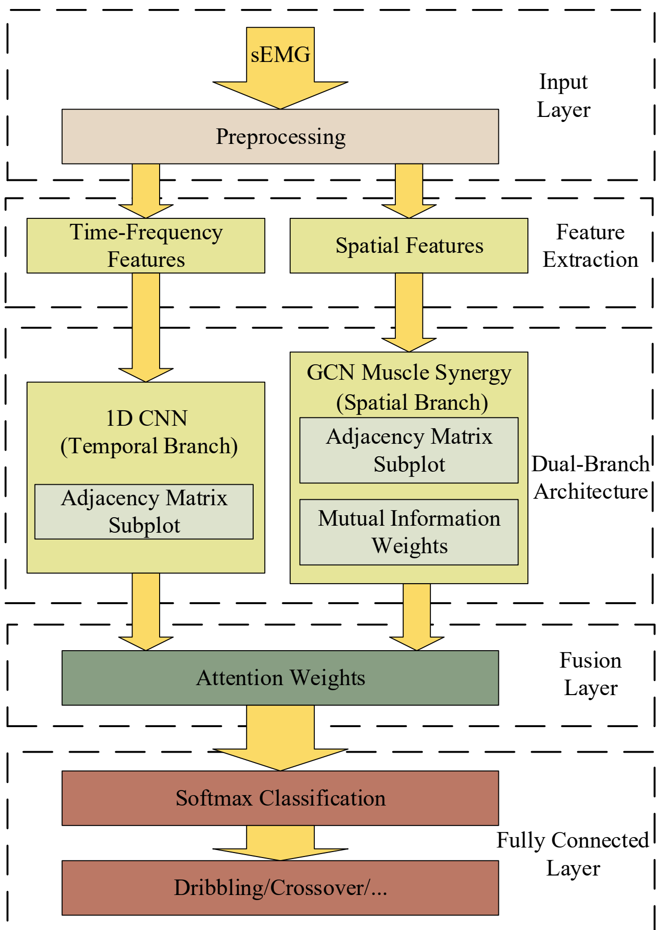
5.2 *Quantification and biomechanical analysis of muscle synergy patterns*

Based on the channel weight matrix W output from the graph convolutional network, the muscle synergy activation map is constructed to reveal the muscle collaboration law of different actions. The muscle synergy index (MSI) is defined as:

$$MSI = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{|W_{ij}|}{\sqrt{W_{ii} W_{jj}}} \tag{29}$$

where N = 8 is the total number of channels, W_{ij} denotes the synergistic strength of channel i^{th} with channel j^{th} is the autocorrelation weight. The closer the MSI value is to 1, the stronger the functional coupling between muscles.

Figure 3 Overall network architecture diagram (see online version for colours)



The reliability of the MSI was validated through multi-dimensional analyses. For professional athletes, the MSI during shooting actions reached 0.89, demonstrating a strong positive correlation with expert technical evaluations averaging 9.2 out of 10 points, as evidenced by a Pearson correlation coefficient of r = 0.86 and p < 0.01. Under fatigue conditions, the MSI for the quadriceps-gastrocnemius muscle pair decreased by 12%, from 0.85 to 0.75, while blood lactate concentration increased significantly from 4.8 to 8.2 mmol/L, yielding a negative correlation of r = –0.79 with p < 0.05. Further analysis highlighted that professional athletes exhibited a significantly shorter activation interval between the deltoid and biceps muscles, at 15±3 ms, compared to amateurs at 45 ± 10 ms, alongside a 30% difference in signal amplitude during critical movement phases. These results confirm the MSI's robustness in quantifying neuromuscular synergy efficiency.

To improve the interpretability of the analysis results, an interactive muscle synergy heat map tool was developed. Coaches can visually assess the quality of the movement through colour shades and side weights, e.g., green highlighted areas indicate highly synergistic muscle groups, while red warning areas indicate imbalances in force generation or potential risk of injury. The tool also compares historical data and generates personalised training recommendations, such as adjusting the jump angle to optimise the lower body power chain.

5.3 Adaptive closed-loop feedback training system

Integrating movement recognition and biomechanical evaluation, the system builds a closed-loop training system of execution-assessment-regulation, which breaks through the limitation of traditional training relying on subjective experience. After the athlete completes the standardised movement wearing the sEMG device, the system calculates the MSI, joint angular velocity (acquired by IMU synchronously) and force timing parameters in real time, and dynamically adjusts the training load based on the PID controller:

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt} \tag{30}$$

where $e(t) = MSI_{target} - MSI_{actual}$ is the synergistic exponential error. $K_p = 0.8$, $K_i = 0.2$, $K_d = 0.05$ are empirical tuning parameters.

The regulation strategy contains two types of mechanisms: positive reinforcement and negative correction. When the athlete has five consecutive MSI $\geq$ 0.9, the system automatically increases the load by 0.5 kg or shortens the shooting distance to increase the training intensity; if the MSI is <0.8 for three consecutive times, the system triggers voice prompts and reduces the training difficulty to avoid the curing of incorrect movements. For injury prevention scenarios, the system will pause training and recommend a rehabilitation programme if unilateral muscle over-activation like rotator cuff injury risk is detected. Measurement data shows that athletes using the system increased their shooting percentage by 19.3% and reduced muscle strain rate by 37.6% within 6 weeks.

6 Experimental results and analyses

The experiment is based on the psublicly available dataset UCI HAR. The dataset contains multichannel sEMG signals from 10 healthy subjects covering five common types of movements. Data were acquired using a Delsys Trigno wireless sensor system with a sampling frequency of 2,000 Hz and electrode attachment locations that strictly followed the SENIAM standard, including key muscle groups such as the biceps, deltoid, and quadriceps. Each type of movement was performed 20 repetitions by subjects under standardised conditions, with a total sample size of 12,000 entries.

In the pre-processing stage, the raw signals were first passed through a 20–500 Hz band-pass filter to eliminate industrial frequency noise and baseline drift, and then segmented with a 200 ms window length and 50 ms overlap to generate time-domain sample segments. Each sample was Z-score normalised to eliminate inter-subject differences in amplitude of muscle activation strength. The experiments were cross-validated using the leave-one-out method. The entire data of one subject at a time was selected as the test set, and the data of the remaining nine were used for training to ensure an objective assessment of the model's generalisation ability. The training phase was done on NVIDIA RTX 3090 GPUs, with the optimiser chosen to be Adam, the initial learning rate set to 0.001 and a cosine decay strategy. The final model is deployed on the Jetson TX2 embedded platform to verify the real-time performance and resource consumption.

In order to comprehensively evaluate the performance of the proposed multi-channel spatio-temporal fusion network (MCFFN), this paper compares five types of mainstream methods, including traditional machine learning methods SVM (Fu et al., 2023), single-channel deep learning models 1D CNN (Ankit et al., 2021) and LSTM (Rezaie et al., 2023) and multi-stream CNN (Wei et al., 2017). Table 1 summarises the average accuracy, F1-score, number of parameters and inference time of each method on the test set.

Table 1 Motion recognition performance comparison (%)

Method	*Accuracy(%)*	*F1-score*	*Number of participants (M)*	*Reasoning time (ms)*
SVM	82.1	80.3	-	2.1
1D CNN	89.5	88.7	2.1	5.7
LSTM	85.2	83.9	1.8	8.3
Multi-stream CNN	91.3	90.1	3.4	10.2
MCFFN	96.2	95.8	4.2	14.7

The experimental results show that MCFFN significantly outperforms other methods with 96.2% accuracy. Compared to traditional SVM, its 14.1% performance improvement validates the advantages of automatic feature extraction for deep learning; the 6.7% gain compared to single-channel CNN stems from the explicit modelling of muscle synergistic relationships by the multi-channel spatio-temporal fusion mechanism. Despite the slightly higher parameter count of MCFFN, its inference time still meets the real-time requirement 30 FPS, indicating that the lightweight design of dynamic deep separable convolution and attention mechanism effectively balances the computational efficiency and accuracy.

Table 2 Results of ablation experiments (average accuracy, %)

Model variants	*Accuracy*	*Performance degradation*
Complete MCFFN	96.2	-
Remove GCN branch	88.9	7.3
Remove attention mechanism	91.6	4.6
Fixed convolution (non-dynamic)	92.4	3.8

To quantify the contribution of each module, we progressively remove key components of the MCFFN for ablation experiments. The result is shown in Table 2.

Experiments show that the absence of the GCN branch leads to a 7.3% performance drop, especially in the change of direction and sharp stop actions where the error rate spikes from 15% to 22%, indicating that explicitly modelling muscle synergy is crucial for complex action recognition. The removal of the attentional mechanism makes it difficult for the model to distinguish the primary and secondary relationships of spatio-temporal features, e.g., the leg noise channel is mistakenly considered as a key feature in shooting actions, leading to a 4.6% decrease in accuracy. After dynamic convolution was replaced with fixed convolution, the model's ability to adapt to signal non-stationarity decreased, and the recognition accuracy of the change-of-direction action was reduced by 6.2%, verifying the necessity of its energy-driven mechanism.

The average accuracy of the model on new subjects who are not involved in training improves from 73.4% to 91.2%, thanks to the fine-tuning of the transfer learning strategy. Specifically, freezing the underlying feature extraction layer, dynamic convolution and GCN, and fine-tuning only the top-level classifier, fully connected layer and attention module, converged in ten iterations with ten sets of target data. Table 3 quantifies the differences in amplitude of key muscle channels before and after migration.

Table 3 Mitigating effect of transfer learning on individual differences (% standard deviation of magnitude)

Muscle channel	*Pre-migration (%)*	*Post-migration (%)*
Biceps	38.2	11.7
Triceps	29.5	9.8
Quadriceps	33.1	12.3
Gastrocnemius	27.5	10.5

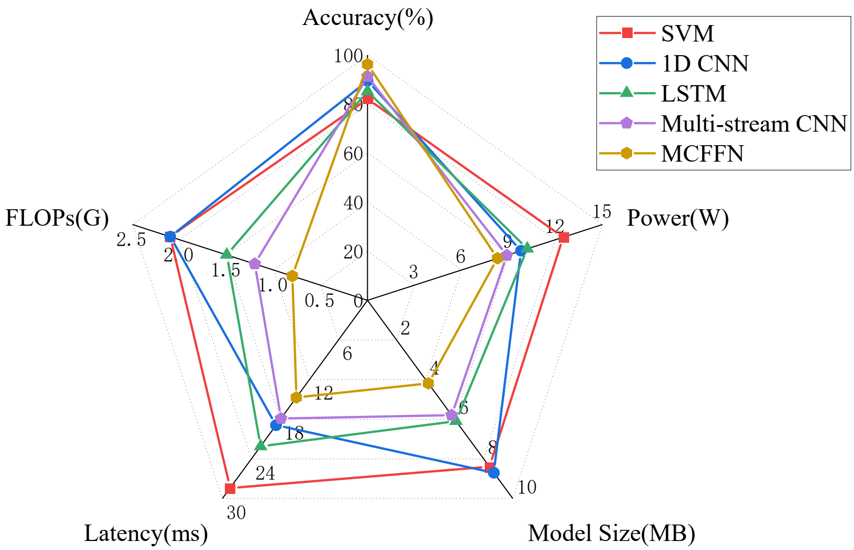
After migration, the inter-individual variation in biceps signal amplitude was reduced from ±38% to ±12%, validating the mitigating effect of the parameter constraint strategy L2 regularisation on distributional bias.

On the Jetson TX2 platform, MCFFN has a single-sample inference time of 14.7 ms and an end-to-end processing latency (including preprocessing) of 48.2 ms, which meets real-time action recognition requirements (>20 FPS). With only 0.8 G floating point operations (FLOPs), the model demonstrates high efficiency suitable for real-time deployment on edge devices. The model occupies a peak memory of 320 MB and a storage size of 4.2 MB, making it suitable for embedded deployments. The power consumption test shows that the average power consumption is 8.3 W for 1 hour of continuous operation, and with a 2,000 mAh battery, it can support 6 hours of continuous monitoring, which meets the needs of court training scenarios. The real-time performance comparison is shown in Figure 4.

Based on the channel weight matrix output from the GCN, the MSI was calculated and compared with the expert assessment results. The MSI for the shooting manoeuvre of professional athletes was 0.89, which was significantly higher than that of the amateur group as 0.72, and was strongly and positively correlated with the biomechanical expert scores on a 10-point scale, with a Pearson correlation coefficient of $r = 0.86$ and $p < 0.01$. Under fatigue conditions, the MSI of the quadriceps and gastrocnemius decreased by 12%, which was negatively correlated with blood lactate concentration from

4.8–8.2 mmol/L. The correlation analysis yielded $r = -0.79$ and $p < 0.05$, confirming the MSI's reliability as a quantitative biomarker for fatigue detection. These findings align with physiological principles, where muscle coordination deteriorates with accumulated metabolic byproducts.

Figure 4 Real-time performance comparison (see online version for colours)



The confusion matrix showed that the model had the highest rate of confusion, between the change of direction and the sharp stop manoeuvre, mainly due to the fact that both relied on rapid knee flexion and extension and had similar sEMG signal morphology. Further analysis revealed that the timing of gastrocnemius activation was delayed by 15 ms for the change-of-direction manoeuvre compared to the sharp stop, while the current model did not explicitly model such subtle differences. Future work could enhance the discriminative nature of the timing features by fusing joint angular velocity data from the inertial measurement unit (IMU). In addition, the absence of adversarial actions such as collision interference in the dataset limits the robustness of the model in real race scenarios, and further expansion of the dataset is needed.

On the whole, the multi-channel spatio-temporal fusion network proposed in this paper demonstrates significant advantages in the basketball action recognition task, and its performance significantly outperforms that of traditional machine learning methods and single-channel deep learning models. By introducing a graph convolutional network to explicitly model the muscle synergy relationship and combining the dynamic attention mechanism to optimise the feature fusion, the model is able to effectively capture the spatio-temporal dependence in complex actions. Ablation experiments further validate the necessity of each key module, with muscle synergy modelling and dynamic feature weighting contributing most significantly to performance improvement. The cross-subject test confirms the generalisation ability of the transfer learning strategy, and the model still maintains high recognition accuracy on data from individuals not involved in training. In addition, the proposed muscle synergy index provides a quantifiable biomechanical basis for movement quality assessment and physiological state monitoring, and its strong correlation with professional scores highlights the interpretability of the model output. The embedded deployment test shows that the system meets the practical application requirements in terms of real-time and energy efficiency, and provides a reliable tool for intelligent analysis of competitive sports and mass fitness.

7 Conclusions

In this paper, a deep learning framework based on spatio-temporal collaborative modelling is proposed to address the multi-channel feature fusion challenges of surface EMG signals in basketball action recognition. By adaptively extracting local muscle activation patterns through dynamic convolution, combining with graph convolutional network to encode muscle functional topology, and designing a cascading attention mechanism to achieve multi-level feature optimisation, the model demonstrates superior performance in complex action classification tasks. The cross-domain transfer learning strategy effectively mitigates the distribution bias problem caused by individual physiological differences, while the introduction of muscle synergy index establishes a data-driven evaluation paradigm for sports technology analysis and injury warning. Experiments demonstrate that the framework not only has high accuracy and strong generalisation, but also its lightweight design can support real-time embedded applications, which provides theoretical support and technical path for the development of intelligent sports equipment.

The current study is based on a generic dataset adapted to basketball actions, and in the future, sEMG data will be collected from real game scenarios to eliminate the semantic gap and improve the model robustness. Future research could further explore multimodal data fusion methods, such as combining inertial sensing and optical capture information, to enhance the model's ability to characterise multidimensional motion features. In the direction of lightweighting, automated techniques such as neural architecture search can be tried to optimise computational efficiency for low-power devices. In addition, expanding the model to other sports scenarios and carrying out long-term clinical validation will help to promote the transformation of the technology from the laboratory to the industry, and ultimately serve the practical needs of athletes' training optimisation and public health management.

Declarations

This work is supported by the Sports Scientific Research Project of Chongqing City Sports Bureau in 2024 named: A Comparative Study on the Difference in Training and Construction of Junior Water Polo Players Reserve Talents in Sichuan and Chongqing Area: A Case Study of Chongqing Water Polo Team's Preparation for the 15th National Games – Strategy, Resources and Effect Analysis (No. A202467).

All authors declare that they have no conflicts of interest.

References

Akira, F., Takuya, I. and Toshio, T. (2021) 'EMG pattern recognition via Bayesian inference with scale mixture-based stochastic generative models', *Expert Systems With Applications*, Vol. 185, p.115644.

Ankit, V., Khimraj, Rajesh, K. and Nilanjan, D. (2021) 'Voting-based 1D CNN model for human lower limb activity recognition using sEMG signal', *Physical and Engineering Sciences in Medicine*, Vol. 44, No. 4, pp.1–13.

Bo, Z., Liu, H. and Song, H. (2021) 'Design and slip prevention control of a multi-sensory anthropomorphic prosthetic hand', *Industrial Robot: the International Journal of Robotics Research and Application*, Vol. 49, No. 2, pp.289–300.

Bryan, L. and Stefan, Z. (2021) 'Time-series forecasting with deep learning: a survey', *Philosophical Transactions of the Royal Society A*, Vol. 379, No. 2194, pp.20200209–20200209.

Ding, H., He, Q., Zeng, L., Zhou, Y., Shen, M. and Dan, G. (2017) 'Motion intent recognition of individual fingers based on mechanomyogram', *Pattern Recognition Letters*, Vol. 88, pp.41–48.

Ding, Z., Yang, C., Wang, Z., Yin, X. and Jiang, F. (2021) 'Online adaptive prediction of human motion intention based on sEMG', *Sensors*, Vol. 21, No. 8, pp.2882–2882.

Feng, L., Zhao, Y., Zhao, W. and Tang, J. (2021) 'A comparative review of graph convolutional networks for human skeleton-based action recognition', *Artificial Intelligence Review*, Vol. 55, No. 5, pp.1–31.

Fu, R., Zhang, B. and Liang, H. (2023) 'Gesture recognition of sEMG signal based on GASF-LDA feature enhancement and adaptive ABC optimized SVM', *Biomedical Signal Processing and Control*, Vol. 85, p.105104.

Ibraheem, M.R. (2023) 'Lower limb analysis based on surface electromyography (sEMG) using different time-frequency representation techniques', *International Journal on Advanced Science, Engineering and Information Technology*, Vol. 13, No. 1, pp.24–33.

Li, X., Yan, L., Zhou, X., Zijian, Y., Lan, T., Peng, F. and Guanglin, L. (2023) 'Simultaneous hand/wrist motion recognition and continuous grasp force estimation based on nonlinear spectral sEMG features for transradial amputees', *Biomedical Signal Processing and Control*, Vol. 85, p.105044.

Mao, W., Ma, B. and Li, Z. (2023) 'STGNN-LMR: a spatial-temporal graph neural network approach based on sEMG lower limb motion recognition', *Journal of Bionic Engineering*, Vol. 21, No. 1, pp.256–269.

Moctar, S.M.S.E., Rida, I. and Boudaoud, S. (2024) 'Comprehensive review of feature extraction techniques for sEMG signal classification: from handcrafted features to deep learning approaches', *IRBM*, Vol. 45, No. 6, pp.100866–100866.

Qi, J., Jiang, G., Li, G., Sun, Y. and Tao, B. (2019) 'Intelligent human-computer interaction based on surface EMG gesture recognition', *IEEE Access*, Vol. 7, pp.61378–61387.

Rezaie, Z.A., Oluwarotimi, W.S., Ali, A., Kianoush, N., A, M.A. and Guanglin, L. (2023) 'An attention-based bidirectional LSTM model for continuous cross-subject estimation of knee joint angle during running from sEMG signals', *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, Vol. 2023, pp.1–4.

Ryu, J. and Kim, D-H. (2017) 'Real-time gait subphase detection using an EMG signal graph matching (ESGM) algorithm based on EMG signals', *Expert Systems with Applications*, Vol. 85, pp.357–365.

Su, Z., Liu, H., Qian, J., Zhang, Z. and Zhang, L. (2021) 'Hand gesture recognition based on sEMG signal and convolutional neural network', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 35, No. 11, p.2151012.

Wei, W., Wong, Y., Du, Y., Hu, Y., Kankanhalli, M. and Geng, W. (2017) 'A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface', *Pattern Recognition Letters*, Vol. 119, pp.131–138.

Xing, K., Yang, P., Huang, J., Wang, Y. and Zhu, Q. (2014) 'A real-time EMG pattern recognition method for virtual myoelectric hand control', *Neurocomputing*, Vol. 136, pp.345–355.

Xiong, B., Chen, W., Niu, Y., Gan, Z., Mao, G. and Xu, Y. (2023) 'A global and local feature fused CNN architecture for the sEMG-based hand gesture recognition', *Computers in Biology and Medicine*, Vol. 166, pp.107497–107497.

Xu, X. and Jiang, H. (2023) 'A hybrid model based on ResNet and GCN for sEMG-based gesture recognition', *Journal of Beijing Institute of Technology*, Vol. 32, No. 2, pp.219–229.

Yasuda, M., Sakata, H., Cho, S., Harada, T., Tanaka, A. and Yokoyama, M. (2017) 'A fast test method for noise robustness of deep neural networks', *IEICE Proceeding Series*, Vol. 29, pp.46–49.

Yogendra, N. (2021) 'Direct comparison of SVM and LR classifier for SEMG signal classification using TFD features', *Materials Today: Proceedings*, Vol. 45, No. P2, pp.3543–3546.

Zhang, Y., Chen, S. and Cao, W. (2021) 'MFFNet: multi-dimensional feature fusion network based on attention mechanism for sEMG analysis to detect muscle fatigue', *Expert Systems with Applications*, Vol. 185, p.115639.

Zhou, W. and Zhou, S. (2021) 'Summary of model and applications of graph convolutional network', *Journal of Research in Science and Engineering*, Vol. 3, No. 1, pp.138–142.