



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

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

Generative adversarial network-enhanced spatio-temporal graph convolution for driving fatigue monitoring in athletic training

Hua Zhao

DOI: [10.1504/IJICT.2026.10077027](https://doi.org/10.1504/IJICT.2026.10077027)

Article History:

Received:	07 September 2025
Last revised:	21 October 2025
Accepted:	22 October 2025
Published online:	26 March 2026

Generative adversarial network-enhanced spatio-temporal graph convolution for driving fatigue monitoring in athletic training

Hua Zhao

Department of Police Physical Training,
Shanxi Police College,
Taiyuan, 030401, China
Email: zhaohua001405@163.com

Abstract: Fatigue monitoring is essential for athletic training. This study addresses class imbalance and scarcity of severe fatigue samples by proposing a spatio-temporal graph convolutional network enhanced with a generative adversarial network. A conditional Wasserstein generative adversarial network generates realistic synthetic skeletal sequences to expand the training set. Combined with spatio-temporal feature extraction, our model achieves end-to-end fatigue level classification. Evaluations on the daily life activities dataset demonstrate superior performance, with accuracy, precision, recall, and F1-score reaching 92.5%, 91.8%, 92.2%, and 92.0% respectively – outperforming support vector machine by 20.2%, long short-term memory by 11.0%, baseline spatio-temporal graph convolutional network by 3.8%, and variational autoencoder-augmented models by 2.8% in accuracy. Ablation studies validate both the generative adversarial network augmentation and nonlinear labelling strategy, offering a reliable vision-based framework for fatigue monitoring.

Keywords: generative adversarial networks; GANs; spatio-temporal graph convolution; exercise fatigue monitoring; data enhancement; Wasserstein generative adversarial networks.

Reference to this paper should be made as follows: Zhao, H. (2026) ‘Generative adversarial network-enhanced spatio-temporal graph convolution for driving fatigue monitoring in athletic training’, *Int. J. Information and Communication Technology*, Vol. 27, No. 28, pp.25–42.

Biographical notes: Hua Zhao received his Master’s degree from the School of Physical Education, Shanxi University, in 2016. He is currently an Associate Professor with the Department of Police Physical Training at Shanxi Police College. His primary research interests include physical education pedagogy, police tactical training, and the application of generative adversarial networks in sports science.

1 Introduction

Exercise fatigue is a natural physiological phenomenon during exercise training characterised by a temporary decrease in the muscle’s ability to produce maximum contraction force or power (Gandevia, 2001). Scientific and accurate monitoring of

fatigue is essential for optimising training loads, enhancing performance, and preventing overtraining and subsequent sports injuries (Enoka and Duchateau, 2008). Traditional fatigue monitoring methods rely heavily on subjective experience or invasive/contact devices. For example, Borg's rating of perceived exertion (RPE), although simple, is highly subjective and subject to individual variation (Borg, 1998). While physiological and biochemical index-based monitoring, such as heart rate variability (HRV) and blood lactate concentration detection, are more objective, they often require wearing specialised sensors or invasive sampling, which interferes with the continuity of the training process and makes it difficult to apply them on a large scale and on a regular basis in natural training scenarios (Achten and Jeukendrup, 2003). Therefore, the development of a non-invasive, contact-free, objective fatigue monitoring technique that can be operated in real time has become a key issue in the field of exercise science and sports engineering that needs to be addressed.

In recent years, the rapid growth of computer vision and deep learning technology has made it possible to analyse human movement without contact using video data. This is a very promising way to solve the problems mentioned above (Cao et al., 2019). In particular, the maturity of deep learning-based pose estimation algorithms, (e.g., OpenPose, AlphaPose) has made it possible to extract the coordinates of two-dimensional (2D) and three-dimensional (3D) skeletal keypoints of the human body in real-time and with high accuracy from monocular red, green, blue (RGB) camera videos (Cao et al., 2019). This has led to a shift in research focus from traditional RGB pixel analysis to the analysis of more semantically informative and compact skeletal sequence data. Skeletal data naturally has high-level structural information and can effectively circumvent the interference of irrelevant factors such as background clutter, lighting changes, and clothing, which lays a solid foundation for in-depth analysis of the intrinsic laws of movement patterns (Zhang et al., 2019). Based on this, a series of research works have attempted to utilise machine learning methods to mine fatigue features from kinematic data. Early methods mostly relied on hand-designed features, (e.g., joint angles, angular velocities, motion smoothness metrics, etc.) and combined them with traditional classifiers such as support vector machines (SVM) or random forest for fatigue recognition (Li et al., 2024). However, the design of manual features relies heavily on expert knowledge and is difficult to capture the subtle and complex patterns of spatio-temporal dynamics under fatigue.

To autonomously acquire more discriminative features, deep learning models, specifically recurrent neural networks (RNNs) and their variants like long short-term memory networks (LSTMs), have been developed to analyse temporal sequences of skeletal keypoints (Liu et al., 2017). These models can somewhat reflect the temporal dependence of activities, but they typically regard the human body as a collection of independent points, neglecting the intrinsic spatial topological linkages among joints. The spatial-temporal graph convolutional network (ST-GCN) approach (Yan et al., 2018) effectively addresses this constraint. ST-GCN creatively represents the human skeleton as a spatio-temporal graph structure, with the nodes representing the joints and the edges representing the physical connections between the bones and the connections that happen over time. ST-GCN can model spatio-temporal characteristics of actions at the same time and in an efficient way by using convolutional operations on the graph structure. It has done very well on large-scale action recognition datasets (Li et al., 2019; Shi et al., 2019). Theoretically, ST-GCN is well suited for capturing fatigue-induced subtle action deformations manifested in the spatio-temporal dimension, such as the decrease in knee

stability during a deep squat or the rise and fall of trunk core stability during a push-up. However, directly applying ST-GCN to fatigue monitoring in athletic training still faces a serious challenge: the scarcity and extreme imbalance of high-quality fatigue data. In real training scenarios, collecting extreme fatigue data from athletes in the exhaustion state is not only difficult, but also carries certain ethical and safety risks (Zhong et al., 2024). Therefore, the number of samples that can be used for training models, especially those labelled as ‘severe fatigue’, is much smaller than that of ‘non-fatigued’ and ‘mildly fatigued’ sample (Jeworutzki, 2025). This data imbalance can easily lead to the training of deep learning models that are biased towards the majority class, making it difficult to accurately identify the most important states of severe fatigue.

As a strong deep generative model, generative adversarial networks (GANs) have a lot of promise for data augmentation and fixing the class imbalance problem (Goodfellow et al., 2014). GANs can learn high-dimensional distributions of real data and make synthetic samples that are very comparable to genuine data through the adversarial game between generators and discriminators (Radford et al., 2015). In areas such as medical image analysis, GANs have been successfully used to generate scarce pathology images to balance the dataset, thereby significantly improving the performance of downstream classification tasks (Frid-Adar et al., 2018). Transferring this to the field of sports fatigue analysis, a natural and innovative idea is: can GANs be utilised to generate human skeletal sequences with a high degree of realism and under different fatigue states? These synthetic data can effectively expand the training set, especially by increasing the diversity of rare fatigue state samples, thus allowing discriminative models such as ST-GCN to learn more robust and generalisable fatigue feature representations. Conditional GAN (Mirza and Osindero, 2014) and Wasserstein GAN (Arjovsky et al., 2017) are two techniques that can be used to create controlled and stable generations. We used them to create temporal skeletal sequences and combine them with the ST-GCN model to monitor sports fatigue. This particular problem is still an under-explored research frontier. This requires the generator to not only produce spatially structured human postures, but also simulate sequences that evolve smoothly over time and conform to the dynamics of a specific fatigue level, which is much more technically challenging than static image generation.

2 Related work

2.1 A vision-based method for monitoring motor fatigue

Automated monitoring of motion fatigue initially relied heavily on wearable sensors, but their invasiveness limited the range of applications. With the development of computer vision technology, vision-based non-contact methods have gradually become a research hotspot. Early studies relied heavily on hand-designed features to quantify changes in motor performance. For example, Li, et al., 2024 investigated the feasibility of vertical F-V curves to monitor the changes in muscle function after fatigue, and its real-time reliability with two fatigue scenarios. 12 males were first tested in deep squatting for one repetition maximum (1RM), and then randomised to participate in heavy loaded training (HLT) and light loaded explosive training (LLB) squatting experiments, with the F-V curves indexed before and after the test. The results showed good reliability of force-velocity (F-V) curves, with both protocols decreasing Pmax, LLB decreasing v0,

and HLT decreasing F0. It was concluded that vertical F-V curves can be used to monitor muscle function and detect the effects of different fatigue protocols. Similarly, Girdhar et al. (2021) in their review systematically summarised attempts to assess muscle fatigue using optical flow, body contouring and kinematic indices of specific body parts. While these approaches have achieved some success, their performance is heavily dependent on the quality of feature engineering, requires deep domain knowledge, and struggles to capture the complex, high-order nonlinear spatio-temporal patterns induced by fatigue. In recent years, deep learning models, particularly convolutional neural networks (CNNs), have been directly used to process raw video frames for automatic feature extraction. However, RGB video-based methods are susceptible to interference from extraneous factors such as background, lighting, and attire, and their models are often complicated and inefficient due to the need to process large amounts of pixel information.

2.2 *Skeletal data-based action recognition with ST-GCN development*

To overcome the limitations of RGB video, the research focus has shifted to behavioural analysis based on skeletal key point data. Skeletal data represents human motion in a more compact and high-level form that naturally provides robustness to background interference. Early methods for analysing skeletal sequences mostly used RNN and their variants such as LSTM. Liu et al. (2017) introduced a global context-aware attention LSTM network aimed at capturing long-range relationships in sequences; it also lends global context to focus on movement-critical joints, and also designs recurrent attention mechanisms to enhance performance. Experiments show that the network can reliably focus on key joints and obtains advanced performance on three relevant datasets. However, RNN-like models typically treat the human body as a series of unorganised points, ignoring the inherent spatial structural relationships between joints. This fundamental limitation was broken by the pioneering work of Yan et al. (2018), who first proposed the ST-GCN. The model depicts the human skeleton as a graph, where nodes signify joints and edges reflect both physical skeletal connections and temporal continuity. ST-GCN can learn features in both spatial configurations and temporal dynamics by designing convolutional processes in both spatial and graph dimensions. Experiments show that ST-GCN significantly improves performance over mainstream methods on the Kinetics and Nanyang Technological University (NTU)-RGBD datasets, with stronger expressive and generalisation capabilities.

Since then, a series of improvement works have emerged to further enhance the performance of ST-GCN. Li, et al. (2019) introduced the action-structure GCN (AS-GCN) to enhance dependency capture through the incorporation of learnable structure graphs and temporal graphs. The network acquires spatio-temporal features for action recognition by utilising action structure graph convolution and temporal sequence convolution as fundamental components. Furthermore, a future posture prediction header is incorporated into the identification header concurrently to obtain more intricate action patterns via self-supervised learning. Validation on two skeleton datasets, NTU-RGB+D and kinetics, demonstrates that AS-GCN markedly enhances performance relative to existing state-of-the-art approaches, and also indicates strong potential for future pose prediction. On the other hand, Zhang et al. (2019) focuses on the view adaptation problem and proposes virtual assistant-RNN (VA-RNN) and VA-CNN, which lets the model handle skeleton data from multiple angles, with a module that determines the optimal viewpoint and transforms the skeleton to aid recognition. Ablation experiments

confirm that the model removes viewpoint effects, and the dual-streaming scheme VA-fusion and random rotation also improve performance and robustness, with the network outperforming the best existing methods on five datasets. These studies fully demonstrate the power of graph-convolutional networks in modelling spatio-temporal features of human movements, and provide a solid theoretical foundation for applying them to tasks that require fine-grained motion analysis, such as fatigue monitoring.

2.3 GANs for data enhancement

Deep learning models need a lot of high-quality training data to work well. In several specific fields, like as medical imaging and fault detection, getting a lot of labelled data is both expensive and hard, and there are typically serious difficulties with class imbalance. Goodfellow et al. (2014) came up with GANs, which are a strong way to solve the challenges above. They do this by learning the distribution of real data through an adversarial game between a generator and a discriminator, which makes it possible to make realistic synthetic examples. The research introduces an innovative framework for adversarial networks to assess the generative model, concurrently training the generative model G (which represents the data distribution) and the discriminant model D (which determines whether the samples are from the training data or from G). G 's training goal is to make D 's error rate as high as possible. The framework is like a two-player minimax game, where there is only one solution for every G and D functions (G cuts down on the training data distribution, while D is always equal to $1/2$). Experiments have confirmed the framework's efficacy when G and D comprise numerous layers of perceptrons, allowing the system to be taught using backpropagation, and eliminating the necessity for Markov chains with approximation inference networks in training or sample generation. Radford et al. (2015) presented the deep convolutional GAN (DCGAN), which uses the CNN as the main part of the GAN architecture. This makes the images that are created much more stable and of higher quality. Deep convolutional adversarial pairs learn a hierarchy of representations, from object parts to scenes, in the generator and discriminator after being trained on a variety of image datasets. The learned features can be used for new tasks, showing that they can be used as a universal image representation.

To achieve control over the attributes of the generated samples, Mirza and Osindero (2014) proposed conditional GAN (cGAN), which guides the generation process by introducing additional conditional information (e.g., class labels). The model implements conditional constraints by providing data y to both the generator and the discriminator, and experiments show that it can generate Modified National Institute of Standards and Technology database (MNIST) numbers conditioned on class labels. In addition, the model can be used to learn multimodal models and also shows potential in image labelling tasks, generating descriptive labels that are not part of the training labels. In the medical domain, Frid-Adar et al. (2018) adeptly utilised GAN to produce synthetic liver lesion images, thereby augmenting the dataset and markedly enhancing the classification model's accuracy. To enhance training stability, Arjovsky et al. (2017) presented Wasserstein GAN (WGAN), which offers smoother gradient signals by substituting Jensen-Shannon (JS) dispersion with Wasserstein distance. Even though GANs have been very successful at making images, their application to temporal data generation, especially the generation of skeletal sequences with complex spatial graph structures, is still at an early stage of exploration. How to utilise GAN to generate sequences that are

both spatially rational and temporally coherent and conform to specific fatigue dynamics patterns is a challenging and cutting-edge problem.

3 Methodology

The GAN-enhanced ST-GCN framework based on GAN enhancement proposed in this paper aims to solve the problem of the scarcity of fatigue samples in real training data by synthesising high-quality fatigue skeleton sequence data, so as to improve the performance of fatigue state monitoring models. The flow of the whole method is shown in Figure 1, which consists of four core stages:

- 1 data pre-processing and fatigue label generation
- 2 ST-GCN design
- 3 GAN enhancement module
- 4 joint training and evaluation.

In the following, we will elaborate on these four parts.

3.1 *Data pre-processing and fatigue label generation*

The publicly available daily life activities (DaLiAc) dataset was used in this study. The dataset contains synchronised video and inertial measurement unit (IMU) data of 19 subjects performing a variety of daily activities, including repetitive movements such as squats and push-ups, which are prone to fatigue induction. The squat and push-up are typical closed-chain compound movements that are prone to fatigue induction, with the following characteristics: large muscle group mobilisation: rapid triggering of observable fatigue state, suitable for data collection; movement standardisation: high repetition, easy to extract fatigue-related features, (e.g., changes in joint angle, decreased stability) from the skeletal sequences; practical application value: it is the core program of strength training, and the results of the study can be directly used for athlete monitoring. The research results can be directly used for athlete monitoring and fitness guidance. The DaLiAc dataset was used in this paper because it contains such movements. We first extracted the 2D skeletal keypoint coordinates of the human body frame by frame from the video sequences using the OpenPose toolbox proposed by Cao et al. (2019). Each human body pose is represented by $N = 25$ joint points, (e.g., neck, left shoulder, right knee, etc.), and the pose of each frame can be denoted as $\mathbf{X}_t \in \mathbb{R}^{N \times C}$. The choice of 25 joint points in this study is not arbitrary, but directly adopts the standard output of OpenPose (BODY_25 model). This model is currently a widely used benchmark tool in the field of skeletal pose estimation, and it strikes a good balance between computational efficiency and descriptive power. 25 joint points cover the major joints of the human body, (e.g., neck, shoulders, elbows, knees, ankles, etc.), which can adequately capture key dynamics in motion, while avoiding the computational redundancy and noise interference caused by too many nodes. This choice ensures the reproducibility of the study and comparability with existing work, and is based on domain consensus rather than ad hoc decision making. Where $C = 2$ represents the 2D coordinate (x, y) channel

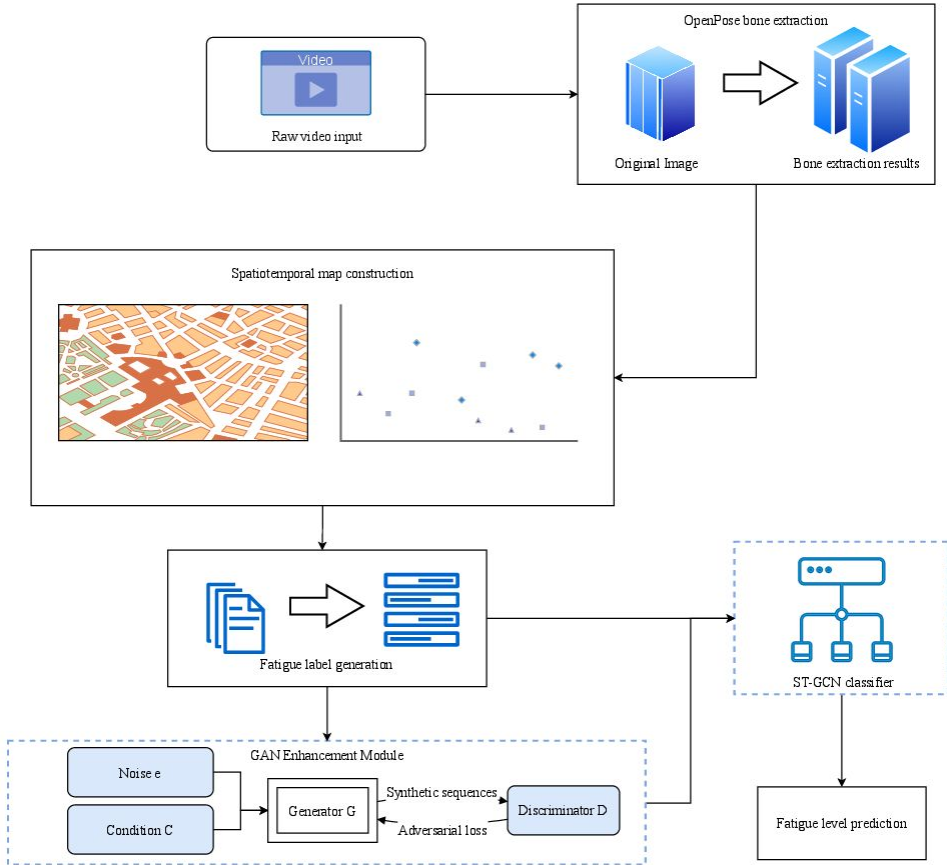
and t is the time index. For a complete action sequence containing T frames, we construct it as a spatio-temporal graph:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \quad (1)$$

where the node set $\mathcal{V} = \{v_{it} \mid t = 1, \dots, T, i = 1, \dots, N\}$ contains all the joint points of all frames. The edge set $\mathcal{E}$ consists of two parts:

- 1 Spatial edges $\mathcal{E}_S$, which connect neighbouring joint points in the same frame, (e.g., the elbow connects the wrist and the shoulder) according to the definition of natural connectivity in the human body.
- 2 Temporal edges $\mathcal{E}_T$, which connect the same joint point in neighbouring frames, thus forming a trajectory in the time dimension.

Figure 1 Schematic diagram of the GAN-enhanced ST-GCN fatigue monitoring framework (see online version for colours)



The DaLiAc dataset provides the exact start and end frames for each movement repetition. Based on this, we assign a fatigue proxy label to each action repetition (rep). Assuming that a subject completes M repetitions consecutively, the fatigue label y_m for

the frame sequence corresponding to the m^{th} repetition is generated by a nonlinear function:

$$y_m = \left(\frac{m}{M} \right)^k \quad (2)$$

where k is a hyperparameter greater than 0 that controls the rate of fatigue growth. When $k = 1$, fatigue grows linearly with repetitions; when $k > 1$, fatigue grows faster in the later stages, which is more consistent with the physiological phenomenon of ‘exhaustion’. The central rationale for the nonlinear labelling strategy (setting the hyperparameter $k > 1$) is the ‘exhaustion effect’ in exercise physiology. The accumulation of fatigue is not a linear process, but accelerates with the number of exercise repetitions. In the later stages of exercise, the accumulation of metabolic wastes and the depletion of energy reserves in the muscles lead to a significant increase in fatigue. By making $k = 2$, the labelling function caused the fatigue value of the late repetitions to increase rapidly, which more realistically simulated this physiological phenomenon. Ablation experiments show that this strategy improves about 2.3% on F1-score than linear labelling ($k = 1$), validating its effectiveness. This continuous $y_m \in (0, 1]$ can be used for the regression task. Alternatively, for classification, we could label the first $\lfloor M/3 \rfloor$ repetitions as ‘no fatigue’ (class 0) and the middle $\lfloor M/3 \rfloor$ repetitions as ‘mild fatigue’ (class 1), Finally, $\lfloor M/3 \rfloor$ times are marked as ‘heavy fatigue’ (class 2).

3.2 ST-GCN fatigue classifier

ST-GCN is the core classifier in this study, whose design follows the seminal work of Yan, et al. (2018) and is adapted for fatigue monitoring tasks. ST-GCN extracts both spatial and temporal features by performing a convolution operation on a constructed spatiotemporal graph. For a given spatio-temporal graph, its graph convolution operation can be defined as information aggregation across both spatial and temporal dimensions.

- Spatial graph convolution: for each node, its features are updated by weighted aggregation of the features of its set of neighbouring nodes $\mathcal{B}(v_{ti})$. The graph convolution operation is defined as:

$$\mathbf{f}_{\text{out}}(v_{ti}) = \sum_{v_{tj} \in \mathcal{B}(v_{ti})} \frac{1}{Z_{ti}(v_{tj})} \mathbf{f}_{\text{in}}(v_{tj}) \cdot \mathbf{w}(l_{ti}(v_{tj})) \quad (3)$$

where $\mathbf{f}_{\text{in}}$ and $\mathbf{f}_{\text{out}}$ are the input and output feature maps. l_{ti} is a partition function that maps each neighbouring node to a unique subset label (e.g., nodes are partitioned into root, centrifugal, and centripetal nodes that are based on how far they are from the skeleton’s centre). $\mathbf{w}$ is a vector of weights based on the partition labels, $l_{ti}(v_{tj})$. $Z_{ti}(v_{tj})$ is the normalisation term for the corresponding subset, which is usually the base of that subset.

- Temporal graph convolution: in the temporal dimension, graph convolution degenerates into performing a standard 1D convolution along the time axis on the trajectories of the same joints. For each node v_{ti} , its temporal neighbours are the same joints $v_{(t+\Delta t)i}$ in the pre Γ frame and the post Γ frame. A learnable convolution kernel Θ is used to perform the convolution operation:

$$\mathbf{f}_{\text{out}}(v_{ti}) = \sum_{\Delta t = -\Gamma}^{\Gamma} \mathbf{f}_{\text{in}}(v_{(t+\Delta t)i}) \cdot \Theta(\Delta t + \Gamma) \quad (4)$$

By alternately stacking multiple spatial map convolutional and temporal convolutional layers, the network is able to progressively extract higher-order, fatigue-related spatio-temporal features from the original joint coordinates. Lastly, a global average pooling layer and a fully connected layer are linked at the conclusion of the network to give the final probability distribution for the fatigue category. The cross-entropy loss function between predicted and true labels is used to learn the parameters of the whole ST-GCN model:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_c} y_{ij} \log(\hat{y}_{ij}) \quad (5)$$

where N_s is the number of training samples, N_c is the number of categories, y_{ij} is the true label (one-hot coding) of sample i belonging to category j , and $\hat{y}_{ij}$ is the model's prediction of the likelihood that sample i belongs to category j .

3.3 GAN data enhancement module

To overcome the problem of scarcity of heavily fatigued samples, we introduce a cGAN to synthesise realistic skeletal sequences. Our GAN architecture is based on the work of Mirza and Osindero (2014) and Arjovsky et al. (2017).

- **Generator (G):** the generator G takes a noise vector from a normal distribution $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ as its input, and a target fatigue condition label $\mathbf{c}$ (after one-hot coding). The input is first spliced and fed into a fully-connected layer, and then reshaped into a series of initial frames. Subsequently, a synthetic skeletal sequence is progressively generated through multiple up-sampling and temporal convolutional layers:

$$\tilde{\mathbf{X}} = G(\mathbf{z} | \mathbf{c}) \quad (6)$$

where $\tilde{\mathbf{X}} \in \mathbb{R}^{T \times N \times C}$. The generator's job is to make sequences that the discriminator cannot tell are true or false.

- **Discriminator (D):** the discriminator D is a binary classifier that takes in a sequence of skeletons $\mathbf{X}$ (which might be actual or fake data) and its conditional label $\mathbf{c}$. It has a number of ST-GCN layers and down-sampling layers. There is a succession of spatio-temporal graphical convolutional and down sampling layers in it. In the end, it gives a scalar $D(\mathbf{X} | \mathbf{c})$, which stands for the chance that the sequence is true given the conditions.
- **Loss function:** to make training more stable, we utilise WGAN-GP loss with a gradient penalty term. You may write the Wasserstein distance like this:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (7)$$

where $\mathbb{P}_r$ is the actual data distribution, $\mathbb{P}_g$ is the created data distribution, and $\Pi(\mathbb{P}_r, \mathbb{P}_g)$ is the collection of all joint distributions.

The generator and discriminator's adversarial loss functions are as follows:

$$\mathcal{L}_D = \mathbb{E}_{\tilde{\mathbf{X}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{X}}|\mathbf{c})] - \mathbb{E}_{\mathbf{X} \sim \mathbb{P}_r} [D(\mathbf{X}|\mathbf{c})] + \lambda \mathbb{E}_{\hat{\mathbf{X}} \sim \mathbb{P}_{\hat{\mathbf{X}}}} \left[\left(\|\nabla_{\hat{\mathbf{X}}} D(\hat{\mathbf{X}}|\mathbf{c})\|_2 - 1 \right)^2 \right] \quad (8)$$

$$\mathcal{L}_G = - \mathbb{E}_{\tilde{\mathbf{X}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{X}}|\mathbf{c})] \quad (9)$$

where $\hat{\mathbf{X}}$ is a point that was sampled by randomly interpolating between the real and produced data, $\mathbb{P}_{\hat{\mathbf{X}}}$ is its distribution, and λ is the gradient penalty coefficient, which is commonly set to 10. We also add a reconstruction loss to the generator to make sure it only makes sequences that meet the conditional labels:

$$\mathcal{L}_{\text{recon}} = \|\mathbf{X} - G(\mathbf{z}|\mathbf{c})\|_2^2 \quad (10)$$

The total loss of the generator is:

$$\mathcal{L}_{G_{\text{total}}} = \mathcal{L}_G + \alpha \mathcal{L}_{\text{recon}} \quad (11)$$

where α is the balancing hyperparameter. The discriminator's job is to make $\mathcal{L}_D$ as small as possible, and the generator's job is to make $\mathcal{L}_{G_{\text{total}}}$ as small as possible.

3.4 Joint training process

We train in two stages. In the first step, we use pre-processed real skeletal sequences and their corresponding fatigue labels to train the cGAN model alone until convergence. In the second step, we use the trained generator G to make a lot of fake skeleton sequences, which are mixed with the real sequences and together form an augmented training set. Finally, we use this augmented dataset to train the final ST-GCN fatigue classifier. The end optimisation target of the whole process can be understood as:

$$\min_{\theta_G} \max_{\theta_D} \mathcal{L}_{G_{\text{total}}}(\theta_G, \theta_D) + \min_{\theta_{\text{ST-GCN}}} \mathcal{L}_{\text{cls}}(\theta_{\text{ST-GCN}}; \mathcal{D}_{\text{real}} \cup \mathcal{D}_{\text{synthetic}}) \quad (12)$$

where θ_G , θ_D , $\theta_{\text{ST-GCN}}$ are the settings for the generator, discriminator, and ST-GCN classifier, respectively, $\mathcal{D}_{\text{real}}$ and $\mathcal{D}_{\text{synthetic}}$ represent the real and synthetic datasets, respectively. In the inference stage, we simply use the trained ST-GCN model to predict the new skeletal sequences, and the generator is no longer involved. This is a design based on module division of labour and efficiency optimisation: the generator is only used for data augmentation in the training phase to solve the sample imbalance problem; the classifier (ST-GCN) is the final deployed inference model, responsible for extracting features from the skeletal sequences and predicting fatigue ratings; the introduction of the generator in inference not only does not help to improve the accuracy, but also increases the computational overhead and system complexity. Therefore, the framework adopts a two-stage decoupled design to ensure efficient and focused inference.

4 Experimental validation

To thoroughly assess the efficacy of the GAN-enhanced ST-GCN framework introduced in this study for fatigue monitoring in athletic training, we designed and conducted exhaustive experiments on the publicly available DaLiAc dataset. This section will first describe the specific setup of the experiments, including data partitioning, metrics for evaluation and baseline models for comparison. Subsequently, we will report the results of the performance comparison with the baseline model, the key ablation experimental analyses, and the visual understanding of the generated data and model decisions to ensure the reliability and interpretability of the experimental conclusions.

4.1 Experimental setup

We closely followed the rules specified by the authors of the DaLiAc dataset and split the data from all 19 participants into three groups: a training set, a validation set, and a test set, with the training set being about 7 times larger than the other two sets. The training set has data from 13 people, whereas the validation and test sets have data from 2 and 4 subjects, respectively. This subject-by-subject strategy ensures that the model evaluates its generalisation ability to unknown individuals, which is in line with practical application scenarios. We mainly employ four generic measures, accuracy, precision, recall, and F1-score (F1), to evaluate performance for the tiredness classification problem in this study. The PyTorch deep learning framework is used to run all of the tests. The Adam optimiser is used with a starting learning rate of 0.001 and a cosine annealing approach to adjust it. Adam integrates Momentum and adaptive learning rate adjustment mechanism, which can maintain an independent learning rate for each parameter, and is suitable for models with non-convex, high-dimensional loss functions such as ST-GCN and GAN. Its advantages include: insensitivity to the initial learning rate, which reduces the difficulty of parameter tuning; adaptive adjustment of the direction of gradient updating, which accelerates convergence; and stable performance in batch training, which is suitable for dealing with skeletal sequence data. The cosine annealing strategy improves the convergence quality of the model by smoothly reducing the learning rate. The learning rate is reduced from the initial value to 0 with the training cycle according to the cosine function, which has the following advantages: keeping the learning rate high at the beginning of the training period to accelerate convergence; slowly decreasing to a very low value at the end of the training period, which is conducive to the model's fine tuning of the parameter and falling into the flat minima, thus improving the generalisation ability; and reducing the oscillations brought by the sudden change of the learning rate compared to the stepwise decrease. This strategy is paired with Adam to further optimise the training process. We set the batch size (batch size) to 32, and all models used early stopping (early stopping) on the validation set to keep from overfitting. Batch size 32 is a trade-off between empirical and computational efficiency in deep learning. Smaller batch sizes (e.g., 16) may lead to large variance in gradient estimation and unstable training; larger batch sizes (e.g., 64) are more stable in terms of gradient, but require more memory and may degrade the generalisation ability. 32 is a widely adopted compromise that efficiently utilises GPU memory while maintaining the stability of training, and is suitable for model training on medium-sized datasets.

To make a fair comparison, we used a number of different baseline models that were typical. The first kind is based on old-fashioned machine learning techniques, where we

used the feature engineering approach described in the work of Li, et al. (2024) to extract a total of 128-dimensional handcrafted features such as the mean velocity of motion, acceleration, and the mean and variance of key joint angles, (e.g., knees, hips) at each joint point from the skeletal sequences, and input them into a linear SVM to perform the classification. The second category is deep learning-based methods, and we implemented the proposed by Zhang et al. (2019) to process the temporal variations of joint points using a two-layer LSTM network, which is a model that can effectively capture the long-term dependencies of movements. The third category is the advanced graph neural network approach, where we replicate the original ST-GCN proposed by Yan et al. (2018), which is the core classifier of our approach, and is used here as a baseline in order to evaluate the effect of GAN enhancement alone. The fourth category is based on other generative models, and we used the variational autoencoder (VAE) proposed by Cemgil et al. (2020) for data enhancement, and constructed a VAE-enhanced ST-GCN baseline to compare the advantages of GAN in generating high-quality spatio-temporal data.

Table 1 Performance comparison of different models on fatigue monitoring tasks

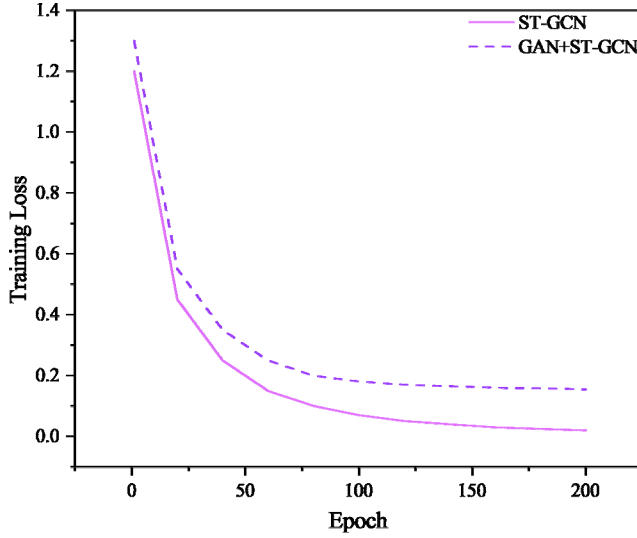
<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
SVM	0.723	0.698	0.710	0.704
LSTM	0.815	0.794	0.802	0.798
ST-GCN	0.887	0.871	0.879	0.875
VAE-ST-GCN	0.901	0.890	0.895	0.892
Proposed (GAN-ST-GCN)	0.925	0.918	0.922	0.920

4.2 Performance comparison and analysis

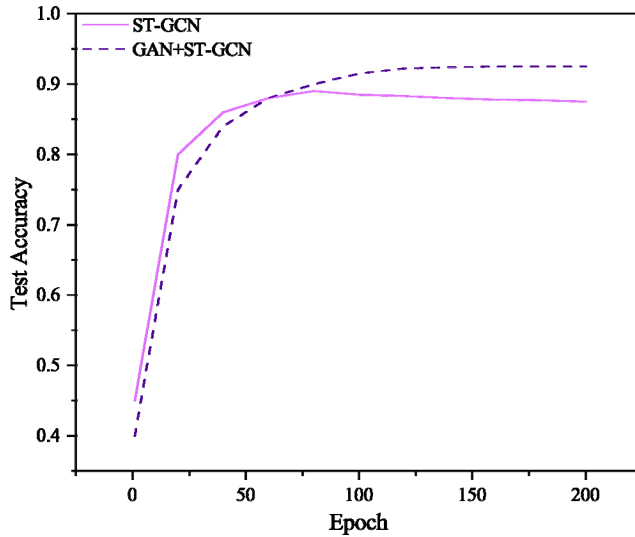
Table 1 shows how well all of the models did on the test set. Our proposed GAN-enhanced ST-GCN framework clearly outperforms all existing baseline models on all assessment measures, with an accuracy of 92.5% and an F1-score of 91.8%. Specifically, the SVM model based on handcrafted features performs the worst, which fully demonstrates that fatigue-related features are extremely complex and subtle and traditional feature engineering relying on expert knowledge is difficult to capture their essence. The LSTM model’s indicators are much better than those of the SVM model. This shows that the deep learning model can learn temporal information on its own quite well, but it ignores the spatial structure information between the joints, and its performance ceiling is obvious. The original ST-GCN model further improves the performance dramatically with an accuracy of 88.7%, which strongly proves the great advantage of graph convolutional networks in modelling spatio-temporal features of human movements simultaneously, which is very suitable for fine-grained movement assessment tasks such as fatigue analysis. Finally, the performance of VAE-enhanced ST-GCN (90.1%), although superior to the original ST-GCN, has limited improvement and is significantly inferior to our approach. We analyse that this is due to the fact that VAE tends to generate samples that are too ‘smooth’ or ‘conservative’, and the diversity and realism of the generated data is not as good as that of the GAN samples obtained through adversarial training, which highlights the superiority of the WGAN in solving

such data generation problems. This highlights the superiority of the WGAN in solving such data generation problems.

Figure 2 Loss function versus accuracy change curve during model training, (a) training loss, (b) test set accuracy (see online version for colours)



(a)



(b)

4.3 Ablation experiment

To thoroughly investigate the role of each fundamental component in the proposed framework, we performed systematic ablation experiments, as illustrated in Figure 2. We first looked at how well the GAN data augmentation module worked. We examined how

well the ST-GCN model trained exclusively on real data worked with the ST-GCN model that also used GAN-generated data. The addition of GAN augmentation made the model more accurate on the test set, going from 88.7% to 92.5% and the F1-score from 87.9% to 91.8%. This indicates that the synthetic skeletal sequences generated by GAN effectively expand the dataset, especially supplementing the sample diversity in the heavy fatigue state, thus significantly improving the generalisation performance and robustness of the final classifier. Second, we verified the impact of the nonlinear coefficient k in the fatigue label generation strategy. We compare the two strategies of linear growth ($k = 1$) and nonlinear growth ($k = 2$). The experimental results show that the strategy with $k = 2$ has an F1-score about 2.3% higher than the strategy with $k = 1$. This is in line with the common sense of exercise physiology that fatigue tends to accumulate at an accelerated rate in the later stages of exercise, and the nonlinear labelling better fits this objective law and provides a more accurate monitoring signal for the model.

4.4 *Visualisation and analytics*

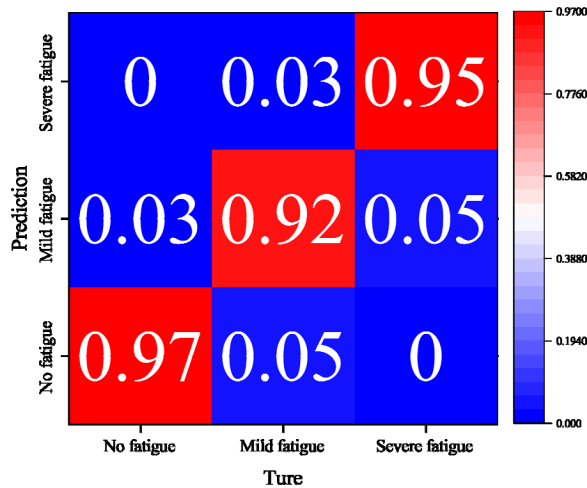
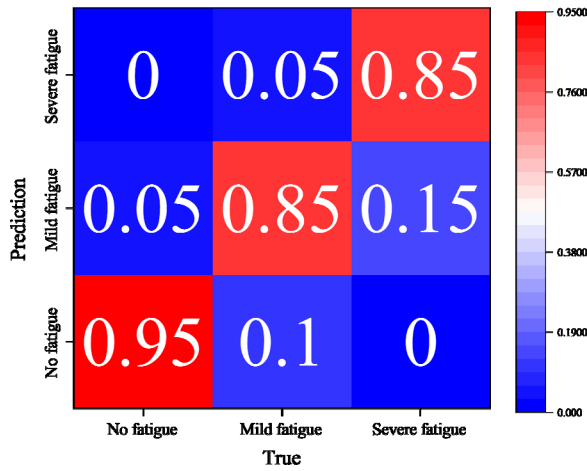
To further understand how well the model classifies things and what specific benefits GAN improvement brings, we performed an in-depth visualisation analysis. First, we made a comparison of the confusion matrices (Figure 3) to show clearly how the original ST-GCN model and our suggested GAN-ST-GCN model classified each tiredness level (no fatigue, mild, and severe) on the test set. The comparison results indicate that the values along the main diagonal of the confusion matrix for the GAN-enhanced model are markedly superior to those of the baseline model, indicating that its overall classification accuracy has been improved; what's more, the misclassification of 'severe fatigue' category, especially the proportion of misclassified as 'mild fatigue', is significantly reduced. More importantly, the misclassification of the 'heavy fatigue' category, especially the proportion of misclassified as 'light fatigue', is significantly reduced, which shows that the GAN's heavy fatigue samples really do help the model recognise the most important categories better., and reduce the risk of underreporting. Second, we visualise the attention weights of the last spatial graph convolution of the ST-GCN classifier when processing a typical squatting movement. It can be clearly observed that in the samples judged to be 'severely fatigued', the model significantly pays more attention to the movement patterns of the lower limb joints, (e.g., knee and hip joints), where swaying and reduced range of motion are typical signs of fatigue. In the 'non-fatigued' sample, attention was more evenly distributed. This visualisation not only corroborates the performance improvements reflected in the confusion matrix, which strengthens the credibility of the model, but also provides interpretable insights for sports scientists that the model has indeed learned to focus on the body part dynamics that are most relevant to fatigue.

4.5 *Experimental results and analysis*

The findings of this work indicate that the integration of GANs with ST-GCNs offers an efficient remedy for the data shortage issue in fatigue monitoring during sports training. Experimental data strongly support our core hypothesis that high-quality and diverse synthetic fatigue skeletal sequences generated by GAN can significantly enhance the generalisation ability and robustness of ST-GCN classification models. The excellent performance of our model on the DaLiAc dataset (92.5% accuracy) is not a coincidence.

Its success can be attributed to two key mechanisms: one, the inherent spatial topology modelling capability of ST-GCN enables it to accurately capture the subtle fatigue-induced degradation manifested in the synergistic movement patterns of multiple joints, which is in the same vein of its strength in fine-grained action recognition as demonstrated by Yan et al. (2018); and two, the introduction of GAN cleverly solves the small-sample learning in the core challenges. As elaborated by Goodfellow et al. (2014), GAN learns the nature of real data distributions through adversarial games, enabling our generators to create synthetic data that are physically plausible and kinetically consistent with fatigue evolution, thus allowing the discriminator, (i.e., the final classifier) to make accurate judgements even in the face of real and rare extreme fatigue states.

Figure 3 Comparison of the confusion matrices of different models on the test set, (a) the original ST-GCN model, (b) the GAN-ST-GCN model proposed in this paper (see online version for colours)



The primary theoretical contribution of this paper is to extend the applicability of deep generative modelling utilising ST-GCN. Previous studies, such as the work of Frid-Adar et al. (2018) in the field of medical imaging, demonstrated the effectiveness of GANs for image data enhancement. We successfully migrate this concept to time-series graph-structured data and validate the great potential of conditional WGANs in generating complex human motion sequences, which provides a new paradigm for solving similar data imbalance problems in the subsequent areas of human behavioural analysis, medical rehabilitation assessment, and so on. Meanwhile, the nonlinear label generation strategy we designed for the specific task of fatigue monitoring also provides learnable experience on how to use domain knowledge to construct effective agent supervision signals in the absence of direct physiological signal labelling.

From a practical perspective, the results of this research have direct application value. It provides coaches and athletes with a non-invasive, low-cost, objective fatigue monitoring tool with real-time feedback. With only an ordinary RGB camera, the system can automatically analyse training movements and assess athletes' fatigue status in real time, thus providing data-driven decision support for scientifically adjusting training loads, optimising training plans, and preventing sports injuries. This is of great significance for both amateurs and high-level athletes.

However, there are several limitations to this study. First, the dataset was acquired in a controlled laboratory environment, which is at odds with real and varied training scenarios. To overcome this limitation, subsequent studies can construct large-scale datasets of real training scenes by deploying a multi-view acquisition system in diverse real training environments, (e.g., gyms, sports fields, outdoors, etc.) to collect dynamic video data containing complex lighting changes, local occlusion, and different backgrounds, and employ domain adaptation and meta-learning methods to enhance the generalisation ability of the model under unknown scenarios. Domain adaptation and meta-learning methods are used to enhance the generalisation ability of the model in unknown scenarios. Second, the fatigue label is a proxy label constructed based on the assumption of the order of motion performance rather than the physiological gold standard of fatigue. For this reason, an improved scheme can introduce multimodal physiological sensing devices, (e.g., surface EMG, heart rate belt, blood lactate metre, etc.) to synchronously collect objective physiological indexes during exercise, and based on the multimodal fusion technique, these physiological gold standard labels can be used to supervise and validate the visual feature model, so as to construct a more accurate and reliable fatigue monitoring system. Third, the current model is a generic model. Third, the current model is a generalised model that does not consider individual differences (e.g., fitness level, technical movement characteristics).

Based on the core findings and technical framework of this study, future research directions should include, first, exploring more efficient generative architectures, (e.g., diffusion models) that will directly improve the data generation capability and robustness of the system. Diffusion modelling is capable of generating more diverse and high-quality skeletal sequence data than traditional GANs. Secondly, the study of cross-modal knowledge distillation methods will significantly improve the accuracy, reliability and credibility of monitoring. Although the current purely visual program is convenient, physiological signals, (e.g., EMG, heart rate) are the 'gold standard' for fatigue. Third, the development of an online learning mechanism will achieve true personalised adaptation and greatly enhance user experience and practicality.

5 Conclusions

In this paper, we provide insights into the problem of contactless motion fatigue monitoring based on computer vision. Aiming at the core challenge of extreme fatigue state samples being scarce and difficult to obtain in this task, we propose a novel GAN-enhanced ST-GCN framework. The framework first utilises OpenPose to extract human skeletal keypoint sequences from raw videos and constructs agent labels of fatigue level based on the action repetition order; subsequently, a conditional Wasserstein GAN is designed to generate realistic synthetic skeletal sequences with specific fatigue features, which effectively expands the training dataset; and finally, a powerful ST-GCN classifier to jointly learn discriminative spatio-temporal fatigue features from raw and synthetic data.

We confirm the superiority of the proposed framework through systematic validation on the publicly available DaLiAc dataset. Experiments show that our approach significantly outperforms traditional machine learning methods, classical LSTM models, and the unenhanced original ST-GCN model on fatigue level classification tasks. A detailed ablation study further reveals the important contributions of each of the GAN data enhancement and nonlinear labelling strategies. The visualisation analysis shows that the generated data are highly consistent with the real data in terms of feature spatial distribution, and the classifiers are able to reasonably focus on the key body parts associated with fatigue.

In summary, this study not only provides a high-performance and practical solution for exercise fatigue monitoring, but more importantly, it provides a proven technological framework to address the challenge of data imbalance in the broader field of human motion analysis. In the future, we will work on deploying and validating the system in real training scenarios and exploring the fusion of multimodal data for more accurate and personalised fatigue monitoring.

Declarations

All authors declare that they have no conflicts of interest.

References

- Achten, J. and Jeukendrup, A.E. (2003) ‘Heart rate monitoring: applications and limitations’, *Sports Medicine*, Vol. 33, No. 7, pp.517–538.
- Arjovsky, M., Chintala, S. and Bottou, L. (2017) ‘Wasserstein generative adversarial networks’, *Machine Learning Research*, Vol. 5, pp.214–223.
- Borg, G. (1998) ‘Borg’s perceived exertion and pain scales’, *Human Kinetics*, Champaign, USA.
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y. (2019) ‘Openpose: realtime multi-person 2d pose estimation using part affinity fields’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 1, pp.172–186.
- Cemgil, T., Ghaisas, S., Dvijotham, K., Goyal, S. and Kohli, P. (2020) ‘The autoencoding variational autoencoder’, *Advances in Neural Information Processing Systems*, Vol. 33, pp.15077–15087.
- Enoka, R.M. and Duchateau, J. (2008) ‘Muscle fatigue: what, why and how it influences muscle function’, *The Journal of Physiology*, Vol. 586, No. 1, pp.11–23.

- Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H. (2018) 'Synthetic data augmentation using GAN for improved liver lesion classification', *International Symposium on Biomedical Imaging*, Vol. 2, pp.289–293.
- Gandevia, S.C. (2001) 'Spinal and supraspinal factors in human muscle fatigue', *Physiological Reviews*, Vol. 81, No. 4, pp.1725–1789.
- Girdhar, P., Johri, P. and Virmani, D. (2021) 'Vision based human activity recognition: a comprehensive review of methods & techniques', *Turkish Journal of Computer and Mathematics Education*, Vol. 12, No. 10, pp.7383–7394.
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) 'Generative adversarial nets', *Advances in Neural Information Processing Systems*, Vol. 27, p.254.
- Jeworutzki, A. (2025) *Strategies for Generalisable Machine Learning with Small Data for Exercise Fatigue Detection*, University of the West of Scotland, Vol. 12, p.443.
- Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y. and Tian, Q. (2019) 'Actional-structural graph convolutional networks for skeleton-based action recognition', *Computer Vision and Pattern Recognition*, Vol. 12, pp.3595–3603.
- Li, Z., Zhi, P., Yuan, Z., Garcia-Ramos, A. and King, M. (2024) 'Feasibility of vertical force-velocity profiles to monitor changes in muscle function following different fatigue protocols', *European Journal of Applied Physiology*, Vol. 124, No. 1, pp.365–374.
- Liu, J., Wang, G., Hu, P., Duan, L.-Y. and Kot, A.C. (2017) 'Global context-aware attention LSTM networks for 3d action recognition', *Computer Vision and Pattern Recognition*, Vol. 20, pp.1647–1656.
- Mirza, M. and Osindero, S. (2014) 'Conditional generative adversarial nets', *Machine Learning*, Vol. 6, p.1784.
- Radford, A., Metz, L. and Chintala, S. (2015) 'Unsupervised representation learning with deep convolutional generative adversarial networks', *Machine Learning*, Vol. 19, p.06434.
- Shi, L., Zhang, Y., Cheng, J. and Lu, H. (2019) 'Skeleton-based action recognition with directed graph neural networks', *Computer Vision and Pattern Recognition*, Vol. 12, pp.7912–7921.
- Yan, S., Xiong, Y. and Lin, D. (2018) 'Spatial temporal graph convolutional networks for skeleton-based action recognition', *Artificial Intelligence*, Vol. 32, No. 1.
- Zhang, P., Lan, C., Xing, J., Zeng, W., Xue, J. and Zheng, N. (2019) 'View adaptive neural networks for high performance skeleton-based human action recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 8, pp.1963–1978.
- Zhong, S., Jia, N., Qu, Y., Zhang, X., Xu, Q., Yang, Y., Wang, Z. and Wang, Z. (2024) 'Analysis and study on biomarkers of local muscle fatigue caused by repetitive lifting task', *BMC Musculoskeletal Disorders*, Vol. 25, No. 1, p.660.