



International Journal of Information and Communication Technology

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

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

Algorithm application and verification for sports training load feature analysis

Jingjing Duan

DOI: [10.1504/IJICT.2026.10077188](https://doi.org/10.1504/IJICT.2026.10077188)

Article History:

Received:	14 October 2025
Last revised:	07 November 2025
Accepted:	11 November 2025
Published online:	01 April 2026

Algorithm application and verification for sports training load feature analysis

Jingjing Duan

School of Hebei Software Institute,
Urban, 071000, China
Email: weipeiyu1981@163.com

Abstract: This study intends to build a hybrid deep learning model based on transformer and bidirectional long short-term memory network (BiLSTM) using high-precision multi-source data of NBA players. First, this study identifies critical load periods and captures sequence global dependence using transformer encoder's multi-head self-attention (MSA). Then, the BiLSTM layer is used to enhance the bidirectional time-series context modelling, and finally the scenario classification results are output. The experimental results demonstrate that in the load trend prediction task, the model's mean squared error (MSE) decreases by 23.3% compared to traditional long short-term memory (LSTM), and by 13.8% compared to the pure transformer model. While the R^2 increases to 0.944, with an improvement of approximately 1.7%, indicating superior fitting accuracy and stability. This model demonstrates significant advantages in both recognition accuracy and generalisation capability. Overall, the proposed method is more sensitive and stable in capturing exercise load's complex temporal dependencies.

Keywords: training load analysis; transformer model; bidirectional long short-term memory network; time series data mining; competitive sports.

Reference to this paper should be made as follows: Duan, J. (2026) 'Algorithm application and verification for sports training load feature analysis', *Int. J. Information and Communication Technology*, Vol. 27, No. 30, pp.73–92.

Biographical notes: Jingjing Duan received her Master's degree in Hebei Normal University. She is currently working in China Hebei Software Vocational and Technical College. Her major fields of study were systematically teaching students various sports skills in accordance with national curriculum standards and school teaching plans, and ensuring teaching quality.

1 Introduction

With the increasingly fierce competition in competitive sports, scientific training has become the core pillar to improve athletes' performance, prolong their sports career and prevent sports injuries (McClean et al., 2024). In this system, the accurate quantification and dynamic analysis of training load play a vital role (West et al., 2024). Training load is usually divided into external load and internal load: the term 'external load' describes the physical strain that athletes endure during practice or competition, including running distance, speed, acceleration, and the frequency of direction changes. Internal load refers

to the corresponding physiological and psychological reactions, such as heart rate, blood lactic acid level and subjective fatigue. Traditionally, coaches mostly rely on empirical observation or a single indicator (such as playing time and total running distance) to evaluate the load. Although these methods are simple and easy, they are difficult to comprehensively and objectively reflect the complex and dynamic load characteristics in competitions and training, especially to effectively capture high-intensity and instantaneous key load events (such as sudden stop and sudden start in continuous defensive rotation), which may bury the hidden dangers of overtraining and sports injury (Vavassori et al., 2024).

Machine learning, especially deep learning algorithm, is being more and more used in sports data analysis because of its powerful ability of automatic feature extraction and complex pattern recognition (Reis et al., 2024). For example, convolutional neural network (CNN) can be used to identify tactical patterns from trajectory data, while recurrent neural network (RNN) and its variants, such as long short-term memory (LSTM), are good at dealing with time sequence dependence and predicting sports performance or fatigue accumulation. However, for the specific task of load feature analysis, the existing methods still face limitations: on the one hand, the traditional LSTM or gated recurrent unit (GRU) model may have the problem of information forgetting or gradient disappearance when dealing with long series, and its one-way structure is difficult to make full use of the complete context information of load events (Wang et al., 2025). On the other hand, although the model based on transformer can effectively capture the global dependence through the self-attention mechanism (SAM), its perception of local time series details is sometimes not as good as that of the special recursive network.

Thus, this study suggests a hybrid deep learning model that combines transformer encoder and bidirectional long short-term memory (BiLSTM) in order to more precisely and finely assess the training load features in basketball games. By utilising massive tracking data provided by the National Basketball Association (NBA) SportVU system and integrating game event labels, this study aims to achieve the following objectives:

- 1 construct a load feature analysis framework capable of integrating global context and local temporal dynamics
- 2 verify the effectiveness of the model in distinguishing load characteristics across different game scenarios (e.g., fast break, half-court offence)
- 3 show the superiority of the suggested model in load trend prediction and fatigue state classification tasks by conducting a methodical comparison with multiple baseline models.

This study expects that the developed algorithm model can provide more scientific decision support for coaching teams, realise personalised monitoring and optimisation of training load, and ultimately contribute to the improvement of athletes' performance and health protection.

Compared with the existing research, the innovations of this study are mainly reflected in three aspects:

- 1 Fusion mechanism innovation: unlike the traditional simple Transformer-LSTM series or independent training schemes, the 'weighted feature concatenation' strategy proposed here achieves adaptive fusion of global and local temporal information

within a unified feature space. This enables the model to establish a more balanced representation between dynamic intensity variations and periodic loads.

- 2 Innovation in data annotation methods: for professional sports field data lacking subjective rating of perceived exertion records, this study proposes an ‘objective fatigue labelling mechanism’ based on individualised physiological recovery rates and short-term performance fluctuations. It achieves effective fatigue classification under conditions devoid of subjective feedback.
- 3 Breakthrough in application scenarios: the research subject of this study is high-frequency, high-intensity data from professional athletes, which fundamentally differs from most studies on exercise load based on laboratory or short-term training environments. This approach can more accurately reflect the physiological-behavioural coupling characteristics under complex competitive conditions, offering direct practical value for competitive sports monitoring and training strategy formulation.

Compared to existing literature that focuses solely on unimodal or single-temporal-scale load modelling methods, this study achieves breakthroughs in algorithm structure, feature fusion, and data annotation mechanisms, offering higher model interpretability and practical application potential.

2 Related work

In the research of sports training load, scholars gradually shift from the traditional single index monitoring to the multi-dimensional evaluation combining internal negative and external loads, and improve the scientificity and operability of monitoring with the help of algorithms and wearable devices. Willberg et al. (2023) focused on the difference between acute and daily training load in traditional basketball and 3×3 basketball match. The local positioning system and inertial measurement unit (IMU) were used to monitor the players. The findings demonstrated that 3×3 basketball players could withstand higher levels of deceleration, acceleration, direction changes, and jumping loads in a given amount of time. Although the average heart rate of the two kinds of competitions was similar, 3×3 basketball was considered to be more fatigued. This showed that in 3×3 basketball, training and tactical design needed to pay more attention to high-speed movement, direction change and projection ability under pressure. From the standpoint of sports medicine, Zhao (2024) discussed about the use of sports biomechanics in the prevention and recovery of sports injuries. He emphasised the core role of kinematics and dynamics analysis in sports performance optimisation and injury prevention, and pointed out the practical value of motion capture system motion capture system (MCS), ground reaction force (GRF) and electromyography (EMG). Future interdisciplinary research will be guided by the article’s emphasis on the possible use of artificial intelligence (AI) and machine learning (ML) in sports load monitoring and injury prevention techniques. In the technical path of monitoring training load, Liu et al. (2024) found that player load (PL) and training impulse (TRIMP) in the lower back position had the highest correlation ($r = 0.843$) by studying the influence of different accelerometer positions on badminton training load monitoring. Furthermore, 88% of the TRIMP variance can be explained by the PL of the lower back and racket player, suggesting that the monitoring technique that

combines internal and external load may greatly increase the accuracy of load measurement. This result provided a practical basis for the layout and algorithm optimisation of wearable devices. Roete et al. (2025) compared the training and competition load differences between ‘very successful’ and ‘not very successful’ seasons based on the long-term season data of professional female cyclists. It was found that in the ‘very successful’ season, athletes generally have a higher total amount of training and competition, but the relative intensity is lower, especially in the power interval 1 and 2. This showed that high-capacity and low-intensity training played a key role in improving endurance and season performance. In contrast, the index of internal load had little difference between different seasons, which highlighted the sensitivity of external load in the evaluation of training effect. The impact of game state on professional football players’ tactical sprint was examined by Lobo-Triviño et al. (2024). The results of the examination of data from Spanish professional league players demonstrated that the position on the field and game condition had a significant impact on tactical sprint performance. For example, when the team wins, the maximum speed of chase action is higher. When the team is behind, the maximum speed of press action is even greater. They emphasised the importance of training design based on competition situation and position difference, which is helpful to realise more personalised load regulation.

Deep learning algorithms have also been extensively studied in this field. Peng and Xu (2023) systematically explored the application of deep learning algorithms in analysing ‘weekly training load’ among students and young athletes. They designed an artificial intelligence model based on a multilayer neural network to model and validate physiological and biochemical indicators throughout the annual training cycle. This model effectively mitigated the tendency of traditional algorithms to fall into local optima through hierarchical feature learning, enabling the identification of the impact of varying training intensities on the musculoskeletal system and neural reactivity. Their findings demonstrated the feasibility of deep learning in modelling nonlinear physiological responses. Tsilimigkras et al. (2024) applied ML methods to predict the risk of non-contact injuries in professional male football players. They comprehensively utilised external load indicators (speed, distance, and acceleration/deceleration) and internal load indicators (heart rate and heart rate recovery rate) to construct an injury risk prediction model based on nine key features. The results showed that the proposed model achieved an accuracy of 0.78, sensitivity of 0.73, and specificity of 0.85 on four-year longitudinal data. It was significantly better than traditional linear regression methods. The systematic review and meta-analysis by Pillitteri et al. (2024) summarised 25 ML-based professional football studies and systematically evaluated the relationship between external load and internal load and their predictive role in injury risk. The results showed a highly positive correlation between external load and internal load indicators, and both types of load indicators were significantly positively correlated with injury risk. Their study also pointed out that most models could more accurately reflect the overall state of training load by integrating the ratio or exponential form of external load and internal load indicators. While ML methods such as random forests and support vector machine (SVM) could automatically identify important features in multidimensional load data and improve injury prediction performance. However, they also pointed out that there were common problems in current research, like limited sample size, insufficient interpretability of algorithms, and weak generalisation ability. Especially when transferring across teams or seasons, the model performance deteriorated significantly.

Based on the above research, it can be seen that ML and deep learning methods have gradually transitioned from feature engineering driven to end-to-end feature learning in the field of motion load analysis, establishing a more refined mapping relationship between load physiological response injury risk. However, existing studies still have several limitations:

- 1 most models rely only on single time scale features and lack comprehensive representation of training load over different time spans
- 2 the interpretability and cross individual generalisation ability of the model are insufficient, and it is difficult to directly migrate to different sports groups.

In response to these issues, this study proposes a Transformer-BiLSTM fusion model that combines global attention and bidirectional temporal modelling in structural design. This model can simultaneously capture the dynamic characteristics of short-term bursts and long-term cumulative loads, thereby improving model interpretability while ensuring prediction accuracy. This provides a more universal and scalable solution for intelligent evaluation of sports loads.

3 Analysis algorithm of physical training load characteristics

3.1 Transformer-BiLSTM

The study suggests a hybrid model that combines a transformer encoder with a BiLSTM network to handle the challenges of capturing contextual information and long-term dependencies in sports load time-series data. Its core components are as follows:

3.1.1 Input layer and positional encoding

The model input consists of preprocessed time-series samples. First, a fully connected linear layer projects the feature vector of each time step to a higher model dimension. Subsequently, to inject sequence order information, sine-cosine positional encoding is introduced and added to the projected features to obtain the final input representation.

3.1.2 BiLSTM module

To further improve its capacity for bidirectional encoding of contextual information, the transformer's output is fed into a bidirectional LSTM layer. Bidirectional sequence encoding: the sequence is processed by the BiLSTM in both forward and reverse directions, producing hidden states for each direction. Each time steps forward and backward hidden states are concatenated. Sequence summarisation: in order to provide a context-aware summary representation of the complete sequence, the forward and reverse hidden states from the BiLSTM's final time step are taken out and concatenated.

3.1.3 Transformer encoder module

The global dependencies between the sequence's various time steps are captured by this module. Mechanism of multi-head self-attention (MSA): an N-layer transformer encoder processes the input. A feed-forward neural network sub-layer and an MSA sub-layer

make up each layer. By employing the self-attention approach, which enables the model to focus on data from every other time step in the sequence at each time step, the model can effectively identify significant movement patterns linked to high load (such as frequent rapid direction changes). Layer normalisation and residual connection: to maintain the stability of the training process, layer normalisation and residual connections are applied after each sub-layer.

3.1.4 Output layer

The output summary vector is transferred to four output neurons, each of which represents one of four load scenario categories: defensive rotation, fast break, half-court offence, and halftime, after it has passed through a fully connected layer with a softmax activation function.

3.1.5 Feature fusion mechanism

Transformer and BiLSTM extract temporal features at both global and local scales, respectively. The number of time steps is T , the output tensor of transformer is $H_T \in \mathbb{R}^{T \times d_T}$, and the output tensor of BiLSTM is $H_B \in \mathbb{R}^{T \times d_B}$. To eliminate the dimensional difference between the two and fuse them in the same representation space, first apply linear mappings (projections) to the two types of outputs separately, as shown in equation (1).

$$\tilde{H}_T = H_T W_T + b_T, \tilde{H}_B = H_B W_B + b_B \quad (1)$$

$W_T \in \mathbb{R}^{d_T \times d_f}$ and $W_B \in \mathbb{R}^{d_B \times d_f}$ are projection matrices, and $b_T, b_B \in \mathbb{R}^{d_f}$ are biases. The mapped uniform dimension $\tilde{H}_T, \tilde{H}_B \in \mathbb{R}^{T \times d_f}$ is obtained. The dimension d_f is set to 256 in the experiment to balance expressive power and computational cost.

After mapping, weighted concatenation is used as the fusion strategy. Let the learnable fusion coefficient be $\in (0, 1)$, and to ensure numerical interval constraints, let $\lambda = \sigma(\alpha)$, where α is the trainable scalar parameter and $\sigma(\cdot)$ is the sigmoid function. It can also be extended to learnable vectors/matrices based on feature dimensions or time steps to obtain finer grained fusion weights. Fusion computing is shown as equation (2).

$$H_F = [\lambda \tilde{H}_T; (1 - \lambda) \tilde{H}_B] \in \mathbb{R}^{T \times 2d_f} \quad (2)$$

The symbol $[\ ; \]$ represents the concatenation operation on the feature dimension (Levi), resulting in a joint representation H_F for each time step. To avoid redundant prediction heads due to excessively large dimensions after concatenation, an additional layer of linear mapping or nonlinear transformation can be applied after concatenation for dimensionality reduction to obtain the final input, as shown in equation (3).

$$H_{out} = \phi(H_F W_o + b_o) \in \mathbb{R}^{T \times d_{out}} \quad (3)$$

$W_o \in \mathbb{R}^{2d_f \times d_{out}}$, $b_o \in \mathbb{R}^{d_{out}}$, are $\phi(\cdot)$ are the activation functions (such as ReLU or a combination of LayerNorm + ReLU), and d_{out} is the feature dimension received by the prediction head (set to 256 in this article). Finally, H_{out} is fed into parallel regression and classification heads to complete load trend prediction and fatigue state classification.

The advantage of this weighted concatenation strategy is that:

- 1 it automatically adjusts the importance of global and local information through the learnable coefficient λ while maintaining the integrity of the two types of representations
- 2 mapping and subsequent dimensionality reduction ensure dimension alignment and parameter controllability
- 3 in end-to-end training, λ is adaptively optimised with task loss to achieve dynamic information balance in different tactical scenarios or individual differences.

3.2 Bi-LSTM time series modelling

In the analysis of sports training load characteristics, athletes' speed, acceleration, heart rate and other parameters often show significant time dependence. In order to capture the dynamic change of load index in time dimension, Bi-LSTM is introduced into the coded multi-dimensional feature sequence (Richter et al., 2024). Bi-LSTM can learn the changing law from the past to the future and the potential trend from the future to the past at the same time, so that the model can understand the time sequence environment of training load more comprehensively.

The forget gate, input gate, and output gate, in turn, regulate the creation of output, the admission of current information, and the retention of historical information in the topology of a conventional LSTM network. They are calculated as equations (4)–(9).

- Forget gate:

$$f_t = \sigma(W_f \cdot [c_{t-1}, h_{t-1}, x_t] + b_f) \quad (4)$$

- Input gate:

$$i_t = \sigma(W_i \cdot [c_{t-1}, h_{t-1}, x_t] + b_i) \quad (5)$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (7)$$

- Output gate:

$$o_t = \sigma(W_o \cdot [c_t, h_{t-1}, x_t] + b_o) \quad (8)$$

- Hidden state:

$$h_t = o_t \cdot \tanh(c_t) \quad (9)$$

x_t denotes input parameters (e.g., speed, heart rate, or acceleration). h_t denotes output parameters. i_t denotes output parameters of the input gate. f_t denotes output parameters of the forget gate. c_t denotes the state of the neuron at time t . o_t denotes output parameters of the output gate. W denotes the weight matrix. b denotes the bias matrix. $\tanh$ and σ denote activation functions (Hu et al., 2025). After the input gate determines which information will be retained, the LSTM stores the current input x_t and the previous state h_{t-1} in the temporary state $\tilde{c}$. By combining the temporary state $\tilde{c}$ and the previous state c_{t-1} , the LSTM unit can obtain the current hidden layer states $\tilde{c}$ and h_t . The forward calculation

of the current LSTM unit is finally completed when they are output and sent to the following LSTM unit. Completing the forward calculation of every LSTM unit along the time axis yields the forward calculation result for the complete network (Ruiye, 2024; Chen et al., 2025). Consequently, the anticipated result at time t is equation (10).

$$a_t = \sigma(h_t \cdot V + z_t) \quad (10)$$

z_t denotes the calculation results of the forget gate, input gate, and output gate of the LSTM at time t before passing through the activation function. V represents the parameter matrix of the hidden layer (Wang et al., 2024). It should also be noted that the calculation methods of the sigmoid activation function and tanh activation function used in the calculation process of the LSTM unit are shown in equations (11)–(12) respectively.

$$\sigma(z_t) = \frac{1}{1 + e^{-z_t}} \quad (11)$$

$$\tanh(z_t) = \frac{e^{z_t} - e^{-z_t}}{e^{z_t} + e^{-z_t}} \quad (12)$$

Unlike unidirectional LSTM, Bi-LSTM establishes both forward and backward dependency paths in the temporal dimension. This makes it possible for the concealed state at each time step to include both possible future load patterns and previous load information (An et al., 2023). This is particularly critical for sports training load prediction. For instance, before an athlete enters the sprint phase, the future variation trends of their heart rate and acceleration hold significant reference value for the judgement of current load (Ahamed and Surputheen, 2022; Jiang, 2022).

Using the function $lstm(x, h, c)$ to integrate the update process of the forward LSTM unit, the transmission mode of the integrated Bi-LSTM are equations (13)–(14).

$$(h_t, c_t) = lstm(x_t, h_{t-1}, c_{t-1}) \quad (13)$$

$$(h_{t-1}^*, c_{t-1}^*) = lstm(x_t, h_t^*, c_t^*) \quad (14)$$

* indicates the information transmitted backward.

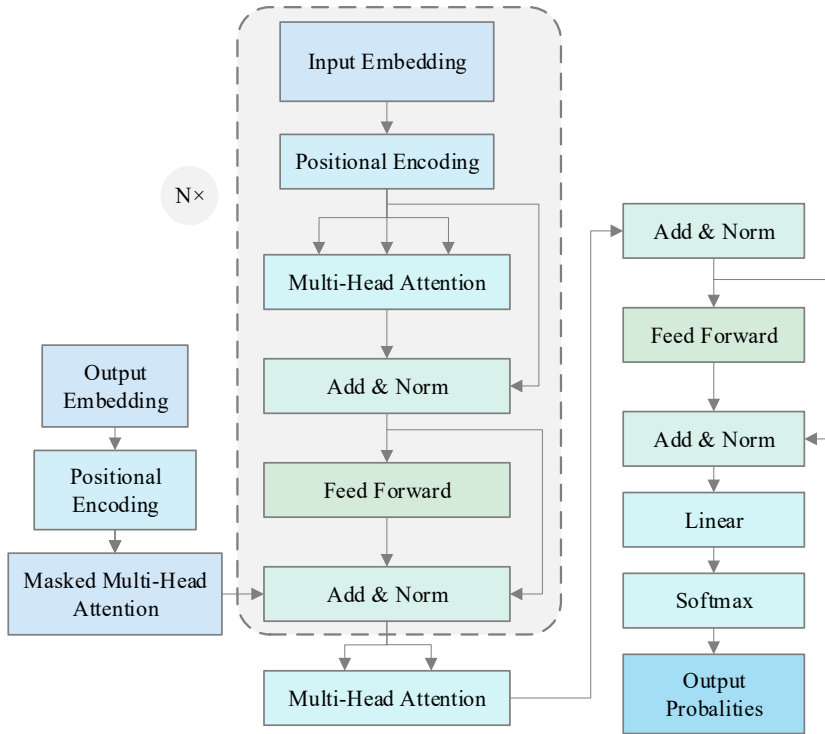
3.3 Transformer model

In the training load characteristic sequence, not all time points and indicators contribute equally to the overall load state. For example, short-term high-intensity sprints, heart rate peaks, or sudden increases in acceleration have a much greater impact on fatigue accumulation than low-intensity stable phases. This study presents an attention mechanism (AM) based on the Bi-LSTM output to allow the model to dynamically focus on these important portions. Its capacity to extract features in parallel greatly enhances training load prediction and fatigue classification performance.

Based on AM, the transformer model's main goal is to model global dependencies using parallel computation. In the scenarios of training load prediction and classification, this study adopts a transformer framework mainly consisting of an encoder structure: the input is a training load sequence $X = (x_1, x_2, \dots, x_n)$. After positional encoding and embedding produce a temporal representation, multi-head attention and feed-forward

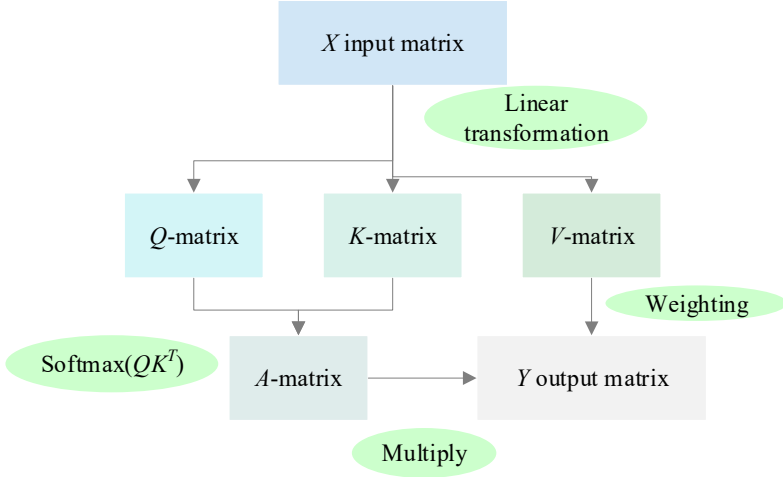
neural network layers are used to extract load features (Chen, 2025). The overall structure is schematically shown in Figure 1.

Figure 1 Schematic diagram of transformer (see online version for colours)



An encoder and a decoder, which are both made up of stacking several similar layers, make up the model shown in Figure 1. The input sequence is processed by the encoder using the feed-forward neural network layer and the MSA mechanism. Decoder then applies a masked MSA mechanism based on this to make sure that subsequent information is hidden while producing outputs. This architecture allows the model to process the entire input sequence in parallel, significantly improving training efficiency (Branquinho et al., 2025). In the figure, it shows that each encoder and decoder layer contains two sub-layers: MSA and feed-forward network. To stabilise the training process and enhance model performance, these sub-layers are followed by addition and normalisation processes, namely residual connections and layer normalisation. To give positional information, positional encoding is provided after the input embedding layer has transformed the input sequence into a vector representation (Chang et al., 2025). Following the conversion of the predicted indices back into word embeddings by the output embedding layer of the decoder, the linear and softmax layers provide the final output probability distribution. The transformer model's MSA mechanism is a crucial component that enables the model to analyse sequences by concentrating on data from various locations.

Figure 2 illustrates the AM's computation procedure.

Figure 2 Computational process of AM (see online version for colours)

The AM's core concept is that the model takes into account both the current input and data from all prior inputs when producing outcomes. Specifically, the model calculates a weight for each element in the input sequence $X = (x_1, x_2, \dots, x_n)$. These weights represent the importance of the element to the current output. The weights are computed through a learnable function, which typically involves interactions between the current element and all elements of the input sequence (Toğaçar et al., 2025).

The core calculation of the AM is divided into two steps:

1 Mapping of queries, keys, and values

First, the input sequence X is projected into query, key, and value vectors, as shown in equation (15).

$$Q = XW^Q, K = XW^K, V = XW^V \quad (15)$$

W^Q , W^K and W^V are learnable weight matrices, respectively.

2 Calculation of attention weight

Calculating the correlation between elements by scaling point product AM, as shown in equation (16).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \quad (16)$$

d_k represents the dimension size of query and key vector. In essence, the calculation process is to quantitatively evaluate the correlation between elements in the input sequence, and then generate a weighted output for each element (Al-Tawil et al., 2025; Xi et al., 2024).

3.4 Model implementation details and training configuration

The implementation details and training configuration of the model are shown in Table 1.

Table 1 Model implementation details and training configuration

<i>Configuration item</i>	<i>Value</i>	<i>Description</i>
Transformer encoder layers N	4	Each layer contains multi head self-attention and feedforward networks.
Multi head attention head count	8	Each dimension is d_model/num_heads.
Model dimension d_model	256	Transformer is consistent with the input projection dimension.
Hidden dimensions in feedforward networks d_ff	512	Feed forward layer width.
BiLSTM hidden unit (per direction)	128	After bidirectional splicing, it becomes 256 dimensions.
Sequence length (window size)	200 time-steps	Align windows with events and form training samples with overlapping slices.
Batch size batch_size	64	Graphics processing unit (GPU) training configuration.
Optimizer	AdamW	weight_decay = 1e-5
Initial learning rate lr	1e-4	Learning rate decay: cosine decay or reduce-on-plateau are optional.
Dropout	0.1	Applied to transformer sub-layers and fully connected layers.
Training rounds (epochs)	60 (early stopping)	Early stopping patience = 8 (stop if there is no improvement in the validation set).
Regression loss	Mean squared error (MSE)	MSE is used for load trend forecasting.
Classification loss	Weighted cross entropy	Adjust weights based on imbalanced categories in the training set.

4 Experimental results and analysis

4.1 Data preprocessing

The dataset used here comes from the NBA SportVU tracking system, covering complete tracking data for some regular season and playoffs of the 2018–2024 season. After data cleaning and window slicing, a total of 42000 time-series samples are formed, of which 70% are used for training, 15% for validation, and 15% for testing. The data types include:

- 1 kinematic characteristics: velocity, step frequency, acceleration, displacement, duration of motion
- 2 physiological characteristics: heart rate distribution range and heart rate recovery and other indicators.

The experimental environment is Ubuntu 22.04, Python 3.10, the deep learning framework is PyTorch 2.2, and the GPU is NVIDIA A100 80 GB.

In the data preprocessing stage, the original time-series data is first subjected to missing and outlier detection. Short-term missing segments (continuous missing time less than 3 sampling points) are smoothed and repaired using linear interpolation, while long-term missing or signal interruptions are directly removed from the corresponding sample window. After identifying extreme outliers (such as instantaneous spikes in acceleration or heart rate) using the quartile range method, replace them with the mean of adjacent sampling points. There are significant differences in numerical scales among different athletes, so all continuous variables are standardised according to the Z-score of the training set statistic. Perform logarithmic transformation and normalisation on variables with highly skewed distribution, such as stride frequency or jump count, to reduce the long tail effect. The time series is then aligned and windowed according to the event timestamp and tactical stage (like fast break, positional attack, defensive transition, etc.), using a sliding window length of 200 frames and a step size of 50 frames for slicing, so that each sample covers the complete local game segment.

The fatigue classification labels are entirely generated based on objective physiological and exercise performance indicators. Firstly, calculate the average performance of each athlete at the individual baseline stage (including average speed, average acceleration, power output proxy, heart rate recovery rate, etc.). Subsequently, calculate the real-time values of the corresponding indicators within each sample window and compare with the athlete's baseline. When the average heart rate recovery (defined as the decrease in heart rate 60 seconds after the competition) of the sample is significantly lower than the 25th percentile of the individual mean, or when the peak acceleration drops by more than the 75th percentile of the individual's long-term distribution, the window is labelled as 'fatigue' (label = 1); otherwise, label it as 'non-fatigue' (label = 0). This dual threshold standard considers the composite characteristics of insufficient physiological recovery and decreased short-term exercise performance, and can objectively classify the training load state without subjective scale conditions. All thresholds are automatically calculated based on training set statistics to avoid manual bias. The final label is confirmed to be reasonable by domain experts through sampling, and during the model training phase, a weighted loss function is used to balance the proportion of fatigue and non-fatigue samples, ensuring that the model can learn stably under imbalanced data conditions.

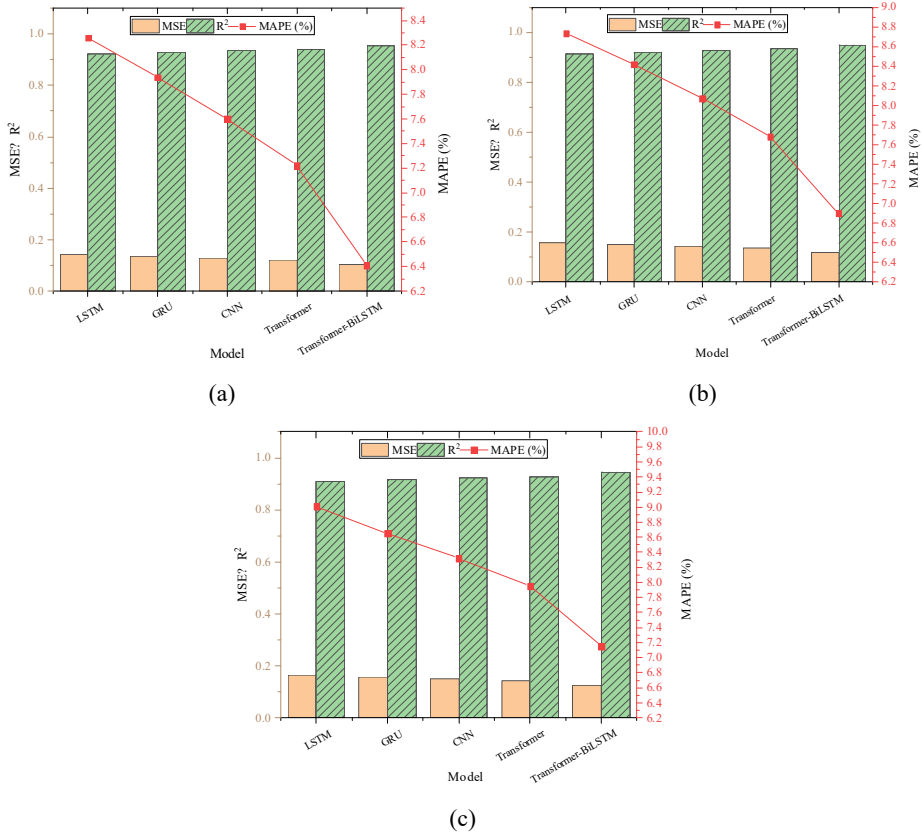
4.2 Performance comparison of training load trend prediction

To fully evaluate the effectiveness of the proposed method, this study divides the data into three subsets: the training set, validation set, and test set. Modelling and performance evaluation are then conducted for training load trend prediction. The evaluation metrics include coefficient of determination (R^2), mean squared error (MSE), and mean absolute percentage error (MAPE).

This study chooses LSTM, GRU, CNN, and pure transformer as deep baselines, and logistic regression (LR), random forest (RF), and SVM as traditional baselines for the following reasons. LSTM/GRU is a typical temporal recursive network, representing the common application of cyclic structures in training load modelling; CNN stands for local pattern extraction method, which is sensitive to short-term signal mutations; transformer represents the latest mainstream architecture for global dependency modelling; LR/RF/SVM represents shallow models based on manual features, which facilitates the demonstration of the superiority of deep end-to-end methods in feature automatic

learning. The reason why some large-scale pre-trained time series models or super-large models generated by automated search have not been included in the current work is that the goal of this study is to demonstrate the interpretability and engineering feasibility of a specific fusion architecture in sports load scenarios. Due to resource and reproducibility considerations, excessive model training costs are high and transfer is complex. The performance comparison of load trend prediction is shown in Figure 3.

Figure 3 Load trend forecasting performance, (a) train (b) validation (c) test (see online version for colours)



From Figure 3(a), all five types of models can effectively fit the load sequence on the training set. Among the traditional recurrent network models, GRU (MSE = 0.136, $R^2 = 0.928$) performs slightly better than LSTM (MSE = 0.142, $R^2 = 0.921$), while CNN further improves the performance to MSE = 0.128 and $R^2 = 0.935$. Transformer shows better performance in global dependency modelling (MSE = 0.121, $R^2 = 0.941$), and the proposed Transformer-BiLSTM model achieves the best results (MSE = 0.104, MAPE = 6.41%, $R^2 = 0.956$), which is significantly superior to the baseline models in both fitting ability and error control.

The results in Figure 3(b) show that the generalisation performance of each model on the validation set is consistent with the trend of the training set. CNN and transformer are superior to LSTM and GRU in error control, but there is still a certain deviation overall.

The Transformer-BiLSTM model still performs the best ($MSE = 0.118$, $MAPE = 6.90\%$, $R^2 = 0.949$). Compared with the standalone transformer model ($MSE = 0.137$, $R^2 = 0.933$), it has significant improvements in both prediction stability and accuracy. This suggests that adding BiLSTM can improve the generalisation impact by effectively enhancing the model's capacity to capture local temporal data.

The performance on the final test set is shown in Figure 3(c). The prediction performance of all models is slightly lower than that in the training and validation phases, but the overall trend remains consistent. The transformer-BiLSTM model still achieves the best results ($MSE = 0.125$, $MAPE = 7.15\%$, $R^2 = 0.944$), which is significantly superior to LSTM ($R^2 = 0.910$) and GRU ($R^2 = 0.915$), and further improves compared with CNN ($R^2 = 0.922$) and transformer ($R^2 = 0.928$). The findings show that while working with unknown data, the model performs better on prediction accuracy and robustness.

Transformer-BiLSTM exhibits notable benefits, particularly in the MSE and MAPE measures, and consistently performs the best across the three phases of the training set, validation set, and test set. This suggests that its fusion technique offers special benefits for modelling both local and global training load features, which can significantly raise trend prediction accuracy and generalisation capacity. It offers dependable algorithmic assistance for sports training load analysis and implementation.

4.3 *Experimental results of fatigue classification*

In the fatigue state recognition task, this study selects three typical ML models as comparison benchmarks, including LR, RF, and SVM, and compares their performance with the proposed Transformer-BiLSTM model. The comparison dimensions cover three indicators: accuracy, F1-score, and AUC to comprehensively evaluate the performance of the classification models. In the fatigue state classification experiment, modelling and comparison are also carried out based on the three subsets of training set, validation set, and test set respectively. Comparison results of the fatigue classification experiment are shown in Figure 4.

In Figure 4(a), on the training set, the Transformer-BiLSTM model achieves the best performance, with accuracy = 0.932, F1-score = 0.917, and AUC = 0.948, which is significantly superior to LR (Accuracy = 0.879, AUC = 0.901), RF (Accuracy = 0.883, AUC = 0.907), and SVM (Accuracy = 0.907, AUC = 0.924). The results indicate that traditional shallow models have limitations in feature extraction and nonlinear relationship modelling. In contrast, the Transformer-BiLSTM model, by combining the global AM and bidirectional temporal modelling, can better capture the complex dynamic features of the fatigue state.

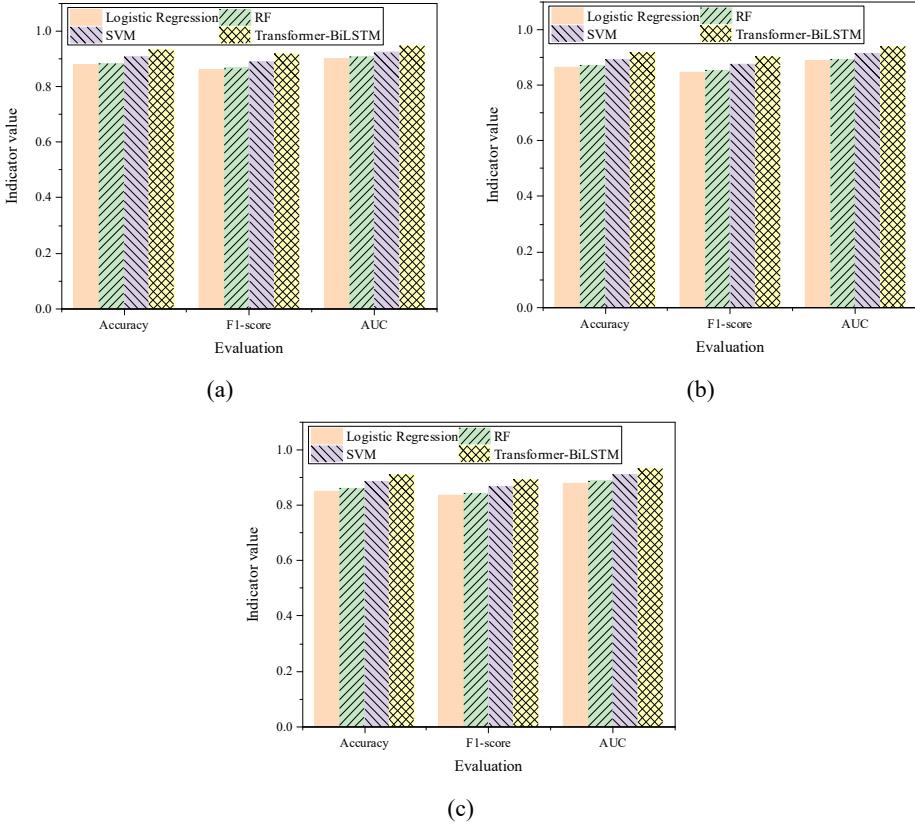
According to Figure 4(b), all models' performance on the validation set somewhat declines, but the general pattern is still in line with that on the training set. The Transformer-BiLSTM still achieves the highest indicators (Accuracy = 0.920, F1-score = 0.903, AUC = 0.940), showing obvious advantages over SVM (accuracy = 0.893, AUC = 0.916). The performance of LR and RF is relatively close, but there is a certain gap between them and the deep learning model. This indicates that the Transformer-BiLSTM has a strong performance in the fitting phase, and has a stable advantage in generalisation ability.

Table 2 Statistical test results and confidence interval range

Task type	Comparison models	Evaluation indicators	The mean of the proposed model $\pm$ SD (standard deviation)	The mean of benchmark model $\pm$ SD	Average increase (%)	t	p	95% confidence interval (lower limit, upper limit)
Load trend forecasting task	LSTM	MSE $\downarrow$	0.125 $\pm$ 0.004	0.163 $\pm$ 0.006	23.3% $\downarrow$	-6.21	<0.001	(-0.054, -0.025)
	Transformer	MSE $\downarrow$	0.125 $\pm$ 0.004	0.145 $\pm$ 0.005	13.8% $\downarrow$	-4.82	0.002	(-0.033, -0.010)
Fatigue state classification task	Transformer	R ² $\uparrow$	0.944 $\pm$ 0.003	0.928 $\pm$ 0.004	+1.7%	5.04	0.002	(0.009, 0.027)
	SVM	Accuracy $\uparrow$	0.911 $\pm$ 0.005	0.886 $\pm$ 0.006	+2.8%	3.94	0.004	(0.010, 0.041)
	SVM	F1-score $\uparrow$	0.895 $\pm$ 0.006	0.869 $\pm$ 0.007	+3.0%	4.27	0.003	(0.012, 0.044)
	SVM	AUC $\uparrow$	0.934 $\pm$ 0.004	0.912 $\pm$ 0.005	+2.4%	3.65	0.006	(0.008, 0.037)

On the final test set [Figure 4(c)], the performance of all models decreases further, but the Transformer-BiLSTM still remains the best, with accuracy = 0.911, F1-score = 0.895, and AUC = 0.934. By comparison, SVM, which is the best among traditional ML models, has accuracy = 0.886 and AUC = 0.912, still lower than those of the Transformer-BiLSTM. This indicates that the proposed model can maintain high classification reliability and discriminative ability when facing unknown data.

Figure 4 Comparison of experimental results of fatigue classification, (a) train (b) validation (c) test (see online version for colours)



In summary, the Transformer-BiLSTM model significantly outperforms traditional shallow learning methods (LR, RF, SVM) in the three core indicators across the training set, validation set, and test set. It performs particularly prominently in the AUC indicator, demonstrating higher sensitivity and robustness for fatigue state classification. In addition to offering a workable technological route for fatigue monitoring and risk early warning in sports training, the finding confirms the efficacy of the suggested approach in tasks involving complicated load feature analysis and fatigue discrimination.

4.4 Statistical significance analysis

To verify whether the performance improvement of the Transformer-BiLSTM model compared to the benchmark model is statistically significant, this section conducts

significance tests and confidence interval analysis on the experimental results. The analysis is based on fivefold cross validation to obtain indicators (MSE, MAPE, R^2 , accuracy, F1-score, and AUC), and independent sample t-test and analysis of variance (ANOVA) are used to evaluate the significance level of differences between each model. Table 2 presents the statistical test results and confidence interval ranges.

As shown in Table 2, the Transformer-BiLSTM model outperforms the benchmark method significantly in all key indicators. In the load trend prediction task, MSE is significantly reduced compared to LSTM ($t = -6.21$, $p < 0.001$), and there is also a significant improvement compared to the transformer model ($t = -4.82$, $p = 0.002$). The improvement in R^2 reaches a statistically significant level ($p < 0.01$), indicating that the model has stronger performance in fitting accuracy and stability. In the fatigue state classification task, the accuracy, F1-score, and AUC of the model improve by 2.8%, 3.0%, and 2.4% respectively compared to SVM, with significant differences ($p < 0.01$). This indicates that the improvement in classification performance and generalisation ability of the model is not caused by random fluctuations. Further one-way ANOVA results show significant differences between model groups for all task types ($F = 9.63$, $p < 0.01$), validating the significant advantage of the proposed model in overall performance.

Furthermore, by calculating the 95% confidence interval, it can be seen that the lower confidence limits for all performance improvements are greater than zero, further demonstrating the statistical robustness of the model performance improvement. Overall, the Transformer-BiLSTM model outperforms various benchmark models in statistical sense, and the performance improvement achieved has repeatability and practical significance, providing more reliable algorithm support for modelling complex motion loads.

4.5 Qualitative case analysis

Take several typical fast attack samples ($n = 120$) and visualise the multi head attention weights and BiLSTM hidden state responses within the model. The results show that within 0–2 seconds of the start of the fast break, the attention weight of the transformer is highly aligned with the peak of rapid acceleration, sudden changes in direction angle, and player near-field confrontation events (the attention peak is synchronised with the peak of acceleration/velocity gradient). Subsequently, BiLSTM shows significant hidden state activation of local patterns before and after the attention highlighting time point (such as speed accumulation before acceleration). This thereby helps the classification head distinguish between ‘high-intensity short-term load in fast attacks’ and ‘low amplitude cumulative load in positional attacks’. This analysis indicates that transformer is responsible for ‘raising suspicion points/keyframes’ (where important load changes may occur) on a global scale, while BiLSTM integrates short-term dynamic information before and after these keyframes into a bidirectional context, thereby forming a more robust judgement of load status. This division mechanism is the intuitive reason why this model is superior to a single model in complex tactical scenarios.

5 Conclusions

This study conducts systematic research based on multi-source time-series data. The main conclusions regarding model construction, experimental verification, and result analysis are as follows:

Firstly, this study constructs a training load characteristic system covering external load, internal load and comprehensive index through multi-dimensional data collected by NBA SportVU system, combined with game logs and some physiological indexes of wearable devices. Among them, the total running distance, average speed and peak acceleration can better reflect the athletes' external physical energy consumption, while the heart rate interval distribution and energy consumption estimation provide an objective quantitative means for the internal physiological load.

Second, methodologically, this study compares a variety of statistical and ML methods, including LR, RF, SVM, CNN, LSTM, GRU, and transformer, and proposes an improved model – Transformer-BiLSTM. Experimental results show that traditional statistical methods have limited performance in trend modelling and are unable to effectively capture complex temporal dependencies. In contrast, deep learning models exhibit obvious advantages in modelling long-term dependencies and feature interactions. In particular, the Transformer-BiLSTM model, by combining the global AM with bidirectional temporal modelling, effectively solves the problems of unidirectional dependency and incomplete information utilisation, achieving significant breakthroughs in both prediction accuracy and classification performance.

Third, the Transformer-BiLSTM model achieves the lowest MSE (0.125) and the highest R2 (0.944) in the training load prediction challenge, demonstrating a considerable improvement over models such as LSTM, GRU, and CNN in experimental verification. In the fatigue classification task, the model achieves an accuracy of 0.911, an F1-score of 0.895, and an AUC of 0.934 on the test set, outperforming traditional methods by roughly 3–6 percentage points. This suggests that the enhanced model may more effectively differentiate between fatigued and non-fatigued states and predict the shifting trend of training load, offering a more useful technical tool for athlete health protection and training optimisation.

The research contribution of this study is as follows: this study does not simply stack transformer and BiLSTM in parallel or series, but makes specialised optimisations for sports training load timing in architecture design and training process. At the input representation level, this study integrates temporal kinematic features with event level semantics (competition event logs), and adopts a cross scale windowing strategy to ensure that the model input contains both short-term high-frequency shocks and medium to long-term cumulative load information. In the encoding layer, this study adopts the order of transformer first and then BiLSTM: transformer is responsible for capturing global dependencies in parallel and highlighting key time points (the attention distribution obtained through multi head self-attention can be directly used for interpretation), and then it is sent to BiLSTM to enhance bidirectional context awareness of local time series, thus taking advantage of BiLSTM's sensitivity to short-term dynamic changes (such as small velocity gradients before and after continuous direction changes). In terms of loss design and multi task training, parallel regression (load trend) and classification (fatigue state) losses are introduced, and a weight adjustment strategy is adopted to balance the impact of short-term shocks and long-term cumulative signals on model parameter updates. The above three designs make the fusion method different

from the general hybrid model in engineering (such as the ‘docking’ realisation of only tandem transformer and LSTM). It is more consistent with the dual attribute of ‘short-term key events + medium and long-term cumulative effects’ of sports load in theory, thus having the unique advantage over single model or blind hybrid model in complex multi-scale training load data.

Lastly, there are important practical ramifications for the research in this study. Professional sports training monitoring systems can directly use the algorithm framework, allowing for the creation of customised training plans during the training process and real-time risk early warning. There are undoubtedly still limitations to this study. For instance, the dataset is limited to basketball, and the generalisation across different sports disciplines requires further verification. Additionally, the model has not yet fully integrated sports psychology and tactical factors. In the future, possibilities of multi-modal data fusion and causal modelling can be explored.

Declarations

The author declares that he has no conflicts of interest.

References

- Ahamed, A.B. and Surputheen, M.M. (2022) ‘Prediction of student’s interest on sports for classification using bi-directional long short term memory model’, *International Journal of Computer Science & Network Security*, Vol. 22, No. 10, pp.246–256.
- Al-Tawil, B., Jung, M., Hempel, T. and Al-Hamadi, A. (2025) ‘Multi-head attention-based framework with residual network for human action recognition’, *Sensors*, Vol. 25, No. 9, p.2930.
- An, Q., Pan, B., Liu, Z., Du, S. and Cui, Y. (2023) ‘Chinese named entity recognition in football based on ALBERT-BiLSTM model’, *Applied Sciences*, Vol. 13, No. 19, p.10814.
- Branquinho, L., Forte, P., De França, E., Ferraz, R., Teixeira, J.E. and Thomatieli-Santos, R. (2025) ‘Training load in sport: current challenges and future perspectives’, *Frontiers in Sports and Active Living*, Vol. 7, No. 1, p.1574500.
- Chang, L., Rani, S. and Akbar, M.A. (2025) ‘ChampionNet: a transformer-enhanced neural architecture search framework for athletic performance prediction and training optimization’, *Discover Computing*, Vol. 28, No. 1, p.71.
- Chen, Y. (2025) ‘Hybrid transformer-LSTM model for athlete performance prediction in sports training management’, *Informatica*, Vol. 49, No. 24, p.342.
- Chen, Z., Xie, Q. and Jiang, W. (2025) ‘Hybrid deep learning models for tennis action recognition: enhancing professional training through CNN-BiLSTM integration’, *Concurrency and Computation: Practice and Experience*, Vol. 37, Nos. 6–8, p.e70029.
- Hu, Y., Li, Y., Cui, B., Su, H. and Zhu, P. (2025) ‘Internet of things enabled deep learning monitoring system for realtime performance metrics and athlete feedback in college sports’, *Scientific Reports*, Vol. 15, No. 1, p.28405.
- Jiang, G. (2022) ‘Construction of correlation analysis model of college students’ sports performance based on convolutional neural network’, *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, p.3621316.
- Liu, T.H., Chen, W.H., Shih, Y., Lin, Y.C., Yu, C. and Shiang, T.Y. (2024) ‘Better position for the wearable sensor to monitor badminton sport training loads’, *Sports Biomechanics*, Vol. 23, No. 4, pp.503–515.

- Lobo-Triviño, D., García-Calvo, T., Rubio-Morales, A., Nevado, F., Chena, M., Piñero-Madrona, J. et al. (2024) 'The influence of match status on the conditional characteristics of tactical sprint actions in professional soccer players', *Biology of Sport*, Vol. 41, No. 2, pp.243–248.
- McClean, Z.J., Pasanen, K., Lun, V., Charest, J., Herzog, W., Werthner, P. et al. (2024) 'A biopsychosocial model for understanding training load, fatigue, and musculoskeletal sport injury in university athletes: a scoping review', *The Journal of Strength & Conditioning Research*, Vol. 38, No. 6, pp.1177–1188.
- Peng, J. and Xu, J. (2023) 'Deep learning analysis on the resulting impacts of weekly load training on students' biological system', *Revista Brasileira de Medicina do Esporte*, Vol. 29, No. 1, p.e2022_0197.
- Pillitteri, G., Petrigna, L., Ficarra, S., Giustino, V., Thomas, E., Rossi, A. et al. (2024) 'Relationship between external and internal load indicators and injury using machine learning in professional soccer: a systematic review and meta-analysis', *Research in Sports Medicine*, Vol. 32, No. 6, pp.902–938.
- Reis, F.J.J., Alaiti, R.K., Vallio, C.S. and Hespanhol, L. (2024) 'Artificial intelligence and machine learning approaches in sports: concepts, applications, challenges, and future perspectives', *Brazilian Journal of Physical Therapy*, Vol. 28, No. 3, p.101083.
- Richter, C., O'Reilly, M. and Delahunt, E. (2024) 'Machine learning in sports science: challenges and opportunities', *Sports Biomechanics*, Vol. 23, No. 8, pp.961–967.
- Roete, A.J., van Erp, T., Elferink-Gemser, M.T., Kempe, M., Lamberts, R.P. et al. (2025) 'Internal and external load characteristics in professional female cyclists during highly successful and less successful seasons', *International Journal of Sports Physiology and Performance*, Vol. 20, No. 9, pp.1224–1231.
- Ruiye, Z. (2024) 'Volleyball training video classification description using the BiLSTM fusion attention mechanism', *Heliyon*, Vol. 10, No. 15, pp.675–681.
- Toğaçar, Ş., Tel, M. and Cömert, Z. (2025) 'Fatigue detection in performance sports with multivariate time series analysis: integrated use of wavelet transform and transformer models', *Signal, Image and Video Processing*, Vol. 19, No. 13, p.1147.
- Tsilimigkras, T., Kakkos, I., Matsopoulos, G.K. and Bogdanis, G.C. (2024) 'Enhancing sports injury risk assessment in soccer through machine learning and training load analysis', *Journal of Sports Science & Medicine*, Vol. 23, No. 3, p.537.
- Vavassori, R., Arroyo, M.P.M. and Espa, A.U. (2024) 'Training load and players' readiness monitoring methods in volleyball: a systematic review', *Kinesiology*, Vol. 56, No. 1, pp.61–77.
- Wang, K., Wang, L. and Sun, J. (2025) 'The data analysis of sports training by ID3 decision tree algorithm and deep learning', *Scientific Reports*, Vol. 15, No. 1, p.15060.
- Wang, K., Zhu, D., Chang, Z. and Wu, Z. (2024) 'Research on prediction and evaluation algorithm of sports athletes performance based on neural network', *Technology and Health Care*, Vol. 32, No. 6, pp.4869–4882.
- West, S., Shrier, I., Impellizzeri, F.M., Clubb, J., Ward, P. and Bullock, G. (2024) 'Training-load management ambiguities and weak logic: creating potential consequences in sport training and performance', *International Journal of Sports Physiology and Performance*, Vol. 20, No. 3, pp.481–484.
- Willberg, C., Wellm, D., Behringer, M. and Zentgraf, K. (2023) 'Analyzing acute, and daily load parameters in match situations – a comparison of classic and 3×3 basketball', *International Journal of Sports Science & Coaching*, Vol. 18, No. 1, pp.207–219.
- Xi, X., Zhang, C., Jia, W. and Jiang, R. (2024) 'Enhancing human pose estimation in sports training: integrating spatiotemporal transformer for improved accuracy and real-time performance', *Alexandria Engineering Journal*, Vol. 109, No. 2, pp.144–156.
- Zhao, F. (2024) 'The application of sports biomechanics in sports injury prevention and rehabilitation', *Frontiers in Sport Research*, Vol. 6, No. 3, pp.142–147.