Digital Intelligence PRESS

Journal of Digital Frontier

https://doi.org/10.67541/jdf2602

Research Article Volume 1, Issue 1 (June 2026) Open Access

Construction of a Student Sports Risk Early Warning Model Based on Multi-Source Data Fusion

Miao Zhou, Huicong Lei *

International Communication and General Education School, Guangxi College of Sports Education, Nanning, Guangxi, China

***Corresponding author:** Huicong Lei. Email: 15578212585@163.com

Abstract

In response to the frequent occurrence of student sports injuries and the lack of real-time warning methods in campus sports scenarios, this paper proposes a student sports risk warning model based on multi-source data fusion. A multi-source acquisition system incorporating wearable physiological signals, video-based human postures, and environmental sensors is constructed, and an adaptive weighted early fusion strategy is proposed to dynamically adjust the contribution of each data source. At the feature extraction level, a multi-scale temporal convolutional network with a multi-head attention mechanism (MHA-MSTCN) and a residual contraction module is designed to capture multi-time-scale risk patterns and suppress motion noise. Furthermore, a dual-path heterogeneous fusion architecture is proposed, using graph attention networks to model joint coupling abnormalities in human motion chains, gated recurrent units to capture temporal drift of physiological parameters, and cross-branch cross-attention to achieve collaborative discrimination of spatial and temporal features. Unsupervised clustering of risk levels is achieved using K-means++ and the silhouette coefficient, and the warning trigger is completed by combining individualized dynamic thresholds and a multi-level delay-tolerant rule engine. Experimental results on a self-built dataset (120 subjects, 53.6 hours of sports data) show that the model has an accuracy of 92.3%, a recall rate of 89.7%, an AUC of 97.1%, a warning lead time of 2.31 seconds, a single-frame inference delay of 7.6 milliseconds, and is significantly superior to baseline models such as LSTM, CNN, and Transformer. Ablation experiments verify the effectiveness of each core module. This model provides a feasible technical solution for intelligent campus sports risk prevention and control.

Keywords

Multi-source data fusion; Sports risk warning; Time-series convolutional network; Graph attention network; Edge computing

Received: 29 May 2026; Revised: 29 Jun 2026; Accepted: 10 Jul 2026; Published: 15 Jul 2026

1. INTRODUCTION

1.1 Research background and significance

Sports injuries among teenagers during school physical activities have become a highly concerning issue in global public health [1]. According to the World Health Organization, sports injuries are among the main types of accidental injuries in school-aged children and adolescents. Approximately one-third of these injuries occur during school physical education classes or after-school exercises. Common sports injuries include ankle sprains, anterior cruciate ligament tears, meniscus injuries, muscle strains, and overuse injuries [2],[3]. These injuries not only cause immediate pain and limited mobility but also may negatively impact students' long-term growth and development, as well as their willingness to participate in sports. Traditional methods of preventing sports risks mainly rely on the experience-based observation and post-event handling by physical education teachers [4],[5]. Teachers need to promptly identify abnormal movements while simultaneously observing the performance of dozens of students, which inherently has limitations in terms of cognitive load and attention allocation. In recent years, with the development of wearable sensor technology, computer vision, and edge computing, data-driven automated sports risk warning has become possible [6]. By integrating multi-source information such as heart rate, acceleration, joint angles, and ground reaction forces, an intelligent model capable of real-time identification of risk precursors can be constructed. This is expected to break through the bottleneck of manual observation and shift the risk prevention and control from occurring after the injury to the early stage of abnormal movements. This research direction holds significant theoretical value and practical significance for improving the safety management level of school sports, reducing the incidence of sports injuries among teenagers, and promoting the development of smart sports education.

1.2 Current research status and limitations

Related scholars have conducted a large number of research studies in the field of movement risk monitoring, which can be roughly classified into three main technical routes. The first type is based on single-modal physiological signals, mainly using indicators such as heart rate variability, blood oxygen saturation, and electromyography to assess exercise load and fatigue levels. For example, some studies predict overtraining syndrome by analyzing the heart rate recovery curve of long-distance runners, while others use the spectral characteristics of surface electromyography to detect early signs of muscle fatigue [7],[8],[9]. The advantage of this type of method lies in the direct correlation between physiological signals and the body's internal state, but its limitation is that it struggles to capture risks at the movement technique level, such as incorrect landing postures or abnormal joint loading patterns, which are often the direct causes of acute sports injuries [10],[11]. The second type is based on computer vision-based posture analysis methods, which extract key points of the human body skeleton using single or multiple cameras and then calculate kinematic parameters such as joint angles, joint angular velocities, and center of mass displacement [12]. This type of method can intuitively reflect the standardization of movement techniques, and typical applications include basketball shooting posture correction and running gait analysis [13],[14],[15]. However, pure visual methods are limited by lighting conditions, occlusion relationships, and the field of view of the camera, and cannot obtain physiological load information of the human body [16],[17]. When students are fatigued but still maintain a seemingly normal surface movement pattern, visual methods struggle to provide early warnings. The third type is based on multi-source heterogeneous data fusion methods, which have gradually become a research hotspot in recent years. Some work attempts to simply concatenate heart rate and acceleration data and input it into traditional machine learning classifiers (such as support vector machines, random forests) for risk classification, and some studies use convolutional neural networks or long short-term memory networks to model the fused time series data [18],[19],[20]. However, most existing fusion methods adopt fixed-weight early fusion or decision-level fusion, lacking the adaptive adjustment ability for the dynamic confidence of data sources. In addition, existing models often use a single time-scale convolution or recurrent structure for feature extraction, while movement risk signals naturally have multi-scale characteristics - heart rate responses belong to slow-changing processes ranging from seconds to tens of seconds, and abnormal joint swings belong to fast-changing processes ranging from milliseconds to hundreds of milliseconds [21],[22]. A single scale is difficult to cover both. In terms of model architecture, most studies treat posture features and physiological features equally and input them into the same network, ignoring the essential differences in spatial structure and temporal evolution between the two: abnormal posture manifests as spatial coupling distortion of multiple joints in the movement chain, while physiological abnormalities manifest as drift and mutation patterns in time series. Homogeneous processing cannot fully exploit the complementary potential of the two types of information [23]. Regarding the warning mechanism, existing work mostly uses fixed thresholds or simple percentile divisions to classify risk levels, failing to consider individual baseline differences and the dynamic change of thresholds during movement, resulting in insufficient individual adaptability and temporal stability of the warnings [24].

1.3 Research contents and innovation points

In response to the aforementioned research limitations, this paper proposes a student sports risk warning model based on multi-source data fusion, which conducts systematic innovations at four levels: data fusion, feature extraction, discriminative architecture, and warning mechanism. At the data fusion level, a multi-source acquisition scheme integrating wearable physiological signals, video-based human postures, and environmental sensors was designed, and an early feature fusion strategy based on local confidence adaptability was proposed. By dynamically calculating the signal quality weights and risk sensitivity weights of each data source at each time step, the model can automatically adjust the contribution ratio of data sources according to the movement stage. At the feature extraction level, the traditional temporal convolutional network was improved, and a multi-head attention multi-scale temporal convolutional network (MHA-MSTCN) was proposed. By deploying four convolutional branches with different dilation rates in parallel to capture risk patterns at different time scales, and by introducing the multi-head attention mechanism to achieve dynamic recombination of multi-scale features, the model can capture the drift and mutation patterns of indicators such as heart rate and blood oxygen. At the discriminative architecture level, an innovative dual-path heterogeneous fusion model was proposed: the first path uses graph attention networks to model the spatial structure of 17 key points of the human movement chain, capturing joint coupling abnormalities through the dynamic attention coefficient between nodes; the second path uses gated recurrent units to model the temporal evolution of physiological parameters, capturing drift and mutation patterns of indicators such as heart rate and blood oxygen. A cross-path cross-attention mechanism was introduced between the two paths to enable spatial abnormal features to query temporal context information, and vice versa, thereby achieving collaborative perception of joint compensation and physiological responses. At the warning mechanism level, an unsupervised clustering method combining K-means++ and the silhouette coefficient was adopted to automatically classify risk levels, and an individualized dynamic threshold adjustment strategy based on historical baselines was proposed. At the same time, a multi-level rule engine with delay tolerance and hysteresis characteristics was designed to effectively balance the sensitivity and false alarm rate of the warning. Based on these innovations, this paper also presents a systematic lightweight design of the model, including sparse graph operations, parameter sharing, temporal cross-attention downsampling, and 8-bit integer quantization, enabling the model to run in real time with a single-frame delay of 7.6 milliseconds on edge computing platforms, thereby meeting the deployment requirements of real-world physical education classes.

1.4 Organization structure of the thesis

This article is divided into seven chapters, organized in a logical sequence from data collection, feature processing, model construction to validation and evaluation. The second chapter elaborates on the collection norms of multi-source motion data, the spatio-temporal alignment and cleaning methods of heterogeneous data, as well as the early feature fusion strategy based on adaptive weighted fusion, providing a unified input representation for subsequent modeling. The third chapter proposes a risk feature extraction module based on the improved temporal convolutional network, including the analysis of the limitations of the basic TCN, the design of multi-head attention multi-scale TCN, the principle of the residual contraction module, and the definition of the feature tensor output of the module. The fourth chapter constructs a dual-path heterogeneous fusion risk discrimination model architecture, introducing the graph attention network branch, the gated recurrent unit branch, the cross-branch cross-attention fusion mechanism, and the lightweight design of the model. The fifth chapter discusses the dynamic division of risk levels and the trigger mechanism of early warning, covering unsupervised clustering based on K-means++ and contour coefficient, individualized dynamic threshold adjustment strategies, and multi-level warning rule engines. The sixth chapter verifies the proposed model through experiments, including the description of the experimental platform and data set, the definition of the evaluation index system, ablation experiments, comparative experiments, robustness and real-time performance tests, and case analysis. The seventh chapter summarizes the research work of the entire article, points out the limitations of the current model, and looks forward to the improvement space in directions such as edge computing adaptation, multi-scenario migration, online incremental learning, and human-computer collaborative interaction.

2. MULTI-SOURCE MOTION DATA ACQUISITION AND PREPROCESSING FUSION FRAMEWORK

2.1 Composition of data sources and acquisition specifications

The multi-source data acquisition system constructed in this study covers the internal physiological state, external posture performance and environmental conditions during the student's movement process. It specifically includes three types of data sources. The first type is wearable physiological signals, obtained through integrated sensor nodes worn on the wrist, chest strap, and ankle, including heart rate, blood oxygen saturation, three-axis acceleration, and skin conductance. The sampling frequency is set at 50 Hz to meet the requirements for detecting movement mutations. The second type is video posture data, which is collected simultaneously from the front side and the rear side by two high-speed RGB cameras. The human 17 key points' two-dimensional coordinate sequences are obtained through the OpenPose skeleton extraction algorithm, with a frame rate of 30 fps [25]. The third type is environmental sensor data, including ground friction sensors, environmental temperature and humidity sensors, with a sampling frequency of 10 Hz. All sensors are calibrated with a unified time synchronization protocol (NTP) before data collection to ensure that the data start time deviation is less than 1 ms [26]. Each student subject continuously collects data during the pre-set six groups of standardized physical exercises (such as shuttle running, standing long jump, sit-ups, etc.), and the on-site sports medicine experts manually mark risk events (such as excessive joint extension, loss of center of gravity, precursors of muscle strain) as the reference benchmark for subsequent modeling.

2.2 Methods for spatio-temporal alignment and cleaning of heterogeneous data

Due to the different sampling frequencies and physical meanings of the three types of data, temporal alignment and cleaning must be performed before they can be integrated. In the time dimension, a unified target sampling frequency of 50 Hz is adopted. For the video posture data, cubic spline interpolation is used to increase the sampling rate from 30 fps to 50 Hz. For the environmental sensor data, a zero-order hold is used to upsample from 10 Hz to 50 Hz, ensuring that each time point has a complete feature vector. In the spatial dimension, the coordinates of the key points of the video skeleton are converted from the pixel coordinate system to the world coordinate system, using the student's hip center when standing as the origin, thereby eliminating the influence of shooting perspective and individual body size differences. The cleaning process uses sliding window outlier detection: for each sensor channel, the local median and absolute deviation of the data within the window are calculated [27]. If a sampling point deviates from the median by more than 3 times the absolute deviation, it is determined as an outlier, and the adjacent valid values are linearly interpolated to replace it. In addition, for data missing segments caused by occlusion or sensor detachment, if the missing length is less than 50 sampling points (i.e., 1 second), a multivariate interpolation method is used for filling, otherwise, the segment is completely removed and masked in the subsequent modeling. After cleaning, each sample is organized as a multi-channel time series, with a length of T time points, and each time point contains D-dimensional features from the three sources.

2.3 Early feature fusion strategy based on adaptive weighted fusion

Early fusion at the data level can maximize the retention of the complementarity of the original information [28]. However, the contribution of different data sources to the risk of movement varies with the type of exercise and time. Therefore, an adaptive weighted fusion strategy is required. Let the multi-source feature vector at time t be $\mathrm{x}_t \in \mathbb{R}^D$, where D is the total feature dimension. In fact, x_t is composed of the physiological feature sub-vector $\mathrm{x}_t^{(p)}$, the posture feature sub-vector $\mathrm{x}_t^{(k)}$, and the environmental feature sub-vector $\mathrm{x}_t^{(e)}$. The traditional weighted fusion uses fixed weights and cannot adapt to dynamic changes. This paper proposes a fusion method based on local confidence adaptive weighting, whose core is to calculate the dynamic weight coefficient that changes with time for each data source.

Define the dynamic weight $w_t^{(s)}$ of the s-th data source ($s \in \{p, k, e\}$) at time t as:

$$w_t^{(s)} = \frac{\exp\left(\alpha \cdot C_t^{(s)} + \beta \cdot S_t^{(s)}\right)}{\sum_{s' \in \{p,k,e\}} \exp\left(\alpha \cdot C_t^{(s')} + \beta \cdot S_t^{(s')}\right)} \tag{1}$$

Where, $C_t^{(s)}$ represents the local signal quality confidence of the data source at time t, defined as the normalized value of the signal-to-noise ratio within the current window, with a range of [0,1]; $S_t^{(s)}$ represents the time sensitivity of the data source for motion risk discrimination, calculated based on the cumulative change rate of the variance of the features in the historical window, with a range of [0,1]; α and β are preset balance coefficients, taking 1.2 and 0.8 respectively, to give slightly higher weight to signal quality. The softmax form of normalization ensures that the sum of the weights of the three sources is always 1, and the differences in quality and sensitivity are amplified through the exponential function. The calculation of the signal quality confidence $C_t^{(s)}$ varies depending on the type of data source: for physiological signals, it uses the ratio of baseline drift amplitude to high-frequency noise power; for posture data, it uses the mean response score of key point detection; for environmental data, it uses the degree of difference between sensor readings and static calibration values.

After obtaining the dynamic weights, the fused feature vector $\tilde{\mathrm{x}}_t$ no longer uses simple weighted averaging, but performs weighted feature mapping, that is, multiplying the feature sub-vectors of different data sources by their respective square root weights and then concatenating them to balance the influence of amplitude:

$$\tilde{x}_t = \left[\sqrt{w_t^{(p)}} \cdot x_t^{(p)}; \sqrt{w_t^{(k)}} \cdot x_t^{(k)}; \sqrt{w_t^{(e)}} \cdot x_t^{(e)}\right] \tag{2}$$

The reason for using square root weights instead of linear weights is that the dimensions of the original feature sub-vectors are different (physiological features usually have lower dimensions, while posture features have higher dimensions). Linear weighting would overly suppress the low-dimensional signals, while square root weights maintain the relative proportions of the features while avoiding extreme scaling. Further, to improve temporal continuity, a Hann window of length 5 is applied to the weight sequence $w_t^{(s)}$ along the time axis to smooth it and prevent sudden changes in weights between adjacent moments from causing oscillations in the fused features [29]. Finally, the fused feature vectors $\tilde{x}_t$ at all times are stacked in chronological order to form the fused multi-source time series data matrix $\widetilde{X} \in \mathbb{R}^{T\times D}$, which serves as the standard input for the subsequent risk feature extraction module. The above adaptive weighted fusion strategy enables the model to automatically adjust the contribution of data sources in different motion stages (such as the posture data being more crucial at the take-off moment and the physiological signals being more important after a long run), providing an earlier fused representation with richer information and controlled noise for subsequent risk discrimination.

3. RISK FEATURE EXTRACTION MODULE BASED ON AN IMPROVED TEMPORAL CONVOLUTIONAL NETWORK

3.1 Basic TCN structure and its limitations in motion risk

The Temporal Convolutional Network (TCN) models the long-term dependencies in time series through causal convolution and dilated convolution, and has advantages such as parallel computing and stable gradients compared to recurrent neural networks [30]. The core component of the basic TCN is the dilated convolution operation. For the input sequence $\widetilde{X} \in \mathbb{R}^{T\times D}$ (T is the time step, D is the dimension of the fused features), the one-dimensional dilated convolution in the time dimension is defined as:

$$F(t) = \sum_{i=0}^{k-1} f(i) \cdot \widetilde{X}(t - d \cdot i) \tag{3}$$

Among them, $F(t)$ represents the output feature value of the convolution kernel at time t, $f(i)$ represents the weight parameter of the convolution kernel at the i-th position, k is the size of the convolution kernel, and d is the dilation rate. $\widetilde{X}(t - d \cdot i)$ represents the vector in the input sequence with the index of $t - d \cdot i$. By stacking multiple layers of dilated convolutions and incrementing the value of d layer by layer, TCN can exponentially expand the receptive field. However, directly applying the basic TCN to the extraction of student sports risk features has three limitations. First, motion risk-related signals have multi-scale characteristics. For example, a heart rate spike may last for several seconds, while joint abnormal oscillations last only milliseconds. A single dilation rate configuration cannot simultaneously capture risk patterns at different time scales. Second, the basic TCN shares the feature extraction weights for each time channel globally, lacking the selective focusing ability on the interdependent relationships between different joints or physiological parameters during the movement process. Precursor risks often manifest in the jointed abnormality in multiple feature dimensions. Third, there is ubiquitous non-stationary noise in sports data caused by action impacts, sensor jitter, etc. The linear convolution operation of the basic TCN will transfer this noise to the deep features, resulting in a decrease in the signal-to-noise ratio of the risk features.

3.2 Improved multi-scale TCN with multi-head attention mechanism (MHA-MSTCN)

In response to the aforementioned limitations, this paper proposes the Multi-Head Attention Multi-Scale Time Series Convolutional Network (Multi-Head Attention Multi-Scale TCN, MHA-MSTCN), whose core idea is to deploy multiple convolutional branches with different dilation rates in parallel, and introduce the multi-head attention mechanism to adaptively fuse multi-scale features. Let the input feature matrix be $\mathrm{H_{in}} \in \mathbb{R}^{T\times D}$. MHA-MSTCN first extracts features at different time scales through four parallel dilated convolutional branches, each using the same kernel size $k = 3$, but with dilation rates $d \in \{1,2,4,8\}$, corresponding to capturing motion change patterns at millisecond, hundred-millisecond, seconds, and tens-of-seconds levels. The output feature map $\mathrm{Z}_j \in \mathbb{R}^{T\times D_{\text{hidden}}}$ of the j-th branch is calculated as:

$$\mathrm{Z}_j = \sigma\left(\mathrm{DilatedConv}_{d_j}(\mathrm{H_{in}})\right) \tag{4}$$

Among them, $\mathrm{DilatedConv}_{d_j}$ represents a one-dimensional convolution operation with dilation rate d_j, $\sigma(\cdot)$ is the ReLU activation function, and D_{hidden} is the dimension of the hidden layer. After obtaining the feature maps of the four branches, they are stacked along the feature dimension to form $\mathrm{Z} = [\mathrm{Z}_1; \mathrm{Z}_2; \mathrm{Z}_3; \mathrm{Z}_4] \in \mathbb{R}^{T\times 4D_{\text{hidden}}}$. Subsequently, a multi-head attention mechanism is introduced to weight and fuse features of different scales, enabling the model to dynamically select the most relevant time scale based on the current motion state. The multi-head attention operation is defined as:

$$\mathrm{Attention(Q, K, V)} = \mathrm{softmax}\left(\frac{\mathrm{QK}^\top}{\sqrt{d_k}}\right)\mathrm{V} \tag{5}$$

Among them, $\mathrm{Q} = \mathrm{ZW}^Q$, $\mathrm{K} = \mathrm{ZW}^K$, $\mathrm{V} = \mathrm{ZW}^V$ represent the query, key and value matrices respectively. $\mathrm{W}^Q, \mathrm{W}^K, \mathrm{W}^V$ are learnable projection matrices, and d_k is the dimension of the key vector. In this study, $h = 8$ attention heads were used for parallel computing, and each head independently learns different feature subspace correlation patterns. Then, the outputs of each head are concatenated and passed through a linear transformation:

$$\mathrm{MultiHead(Z)} = \mathrm{Concat}(\mathrm{head}_1, \dots, \mathrm{head}_h)\mathrm{W}^O \tag{6}$$

Among them, $\mathrm{head}_i = \mathrm{Attention}(\mathrm{ZW}_i^Q, \mathrm{ZW}_i^K, \mathrm{ZW}_i^V)$, and $\mathrm{W}^O \in \mathbb{R}^{hd_v\times 4D_{\text{hidden}}}$ is the output projection matrix. The multi-head attention mechanism enables the model to establish a global receptive field in the time step dimension, and can nonlinearly recombine features from different time scales, effectively solving the problems in the basic TCN where the single expansion rate information is incomplete and there is a lack of selective focusing between channels. After being weighted by multi-head attention, it is further added to the original input through a residual connection and executed with layer normalization, resulting in multi-scale enhanced features $\mathrm{H_{ms}} \in \mathbb{R}^{T\times 4D_{\text{hidden}}}$.

3.3 Residual shrinkage module suppresses motion noise interference

The non-stationary noise generated during the movement (such as the accelerometer saturation noise caused by landing impact, and the skeletal jitter noise during rapid turns) can significantly interfere with the quality of risk features. Traditional methods usually adopt a fixed low-pass filter, but they will also filter out high-frequency risk signals (such as the precursors of muscle tremors). In this paper, a residual contraction module is embedded after MHA-MSTCN. This module adaptively eliminates noise components irrelevant to risk discrimination through soft thresholding. The core operation of the residual contraction module is to perform thresholding contraction on the input feature H_{ms}:

$$\mathrm{H}_{\mathrm{denoise}} = \mathrm{sign}(\mathrm{H}_{\mathrm{ms}}) \cdot \max(|\, \mathrm{H}_{\mathrm{ms}} \,| - \tau, 0) \tag{7}$$

Here, $\mathrm{sign}(\cdot)$ is the sign function, $|\cdot|$ represents element-wise absolute value, and τ is a learnable threshold parameter. This operation sets feature components with amplitudes less than the threshold to zero (considered noise), retains those with amplitudes greater than the threshold (considered valid signals), and maintains the differentiability of the function to support end-to-end training. The threshold τ is not a fixed value but is dynamically generated based on the features themselves. Specifically, the H_{ms} is globally averaged pooled in the channel dimension to obtain the channel description vector, and then it passes through two fully connected layers and the Sigmoid activation function to output a scaling coefficient. This coefficient is multiplied by the average of the absolute values of the features to obtain the adaptive threshold for each sample:

$$\tau = \mathrm{Sigmoid}(\mathrm{W}_2 \cdot \mathrm{ReLU}(\mathrm{W}_1 \cdot \mathrm{GAP}(\mathrm{H}_{\mathrm{ms}}))) \cdot \mathrm{mean}(\ |\, \mathrm{H}_{\mathrm{ms}} \,|\ ) \tag{8}$$

Among them, $\mathrm{GAP}(\cdot)$ represents the global average pooling operation, and W_1 and W_2 are the weight matrices of the fully connected layer. This adaptive threshold mechanism can dynamically adjust the intensity of noise suppression according to different exercise intensities and individual differences among students: in intense exercise, the threshold automatically increases to filter out the impact noise, and in stable exercise, the threshold automatically decreases to retain the subtle physiological warning signals. After soft thresholding is completed, a 1×1 convolution layer is used to map the feature dimensions back to D_{hidden} and form a residual connection with the module input to ensure smooth gradient transmission.

3.4 Module output: high-dimensional risk feature tensor

After the above multi-scale temporal convolution, multi-head attention fusion, and residual contraction denoising processes, the original multi-source fusion data is transformed into an information-rich and noise-suppressed risk representation. Let the length of the input time window be T (in the experiment, T = 256, corresponding to approximately 5 seconds of motion data), and the final high-dimensional risk feature tensor output by the module is denoted as $\mathrm{R} \in \mathbb{R}^{T \times D_{\mathrm{risk}}}$, where the dimension D_{risk} of the risk features is set to 128. Each time slice $\mathrm{R}_t \in \mathbb{R}^{D_{\mathrm{risk}}}$ of this tensor encodes the multi-scale motion risk information within the neighborhood at the t-th time, and due to the effect of the residual contraction module, the noise channels are effectively suppressed. Moreover, to preserve the temporal sequence for subsequent risk evolution modeling, the output tensor R maintains the same time steps T as the input and does not undergo any downsampling operation. This high-dimensional risk feature tensor will be passed to the dual-path heterogeneous fusion risk discrimination model described in the next chapter as the basic risk representation at its input end. It is worth emphasizing that the combination of MHA-MSTCN and the residual contraction module enables the network to simultaneously complete three tasks: multi-scale temporal mode extraction, cross-scale feature dynamic fusion, and noise adaptive elimination independent of the motion task, thereby providing a highly discriminative and robust feature basis for subsequent risk discrimination [31].

4. ARCHITECTURE OF THE DUAL-PATH HETEROGENEOUS FUSION RISK DISCRIMINATION MODEL

4.1 Branch 1: graph attention network models the abnormal coupling of joints in the human motion chain

During human movement, risks often do not occur independently in a single joint, but manifest as abnormal coupling between multiple joints in the movement chain. For example, excessive inward flexion of the knee joint is often accompanied by external rotation of the ankle joint and compensatory rotation of the hip joint. To capture this spatial structural abnormality, the branch employs the Graph Attention Network (GAT) to model the coupling relationship between the joints of the human skeleton. Firstly, each moment in the high-dimensional risk feature tensor $\mathrm{R} \in \mathbb{R}^{T\times D_{\text{risk}}}$ is decomposed onto 17 key points of the human body, and a joint feature matrix $\mathrm{J}_t \in \mathbb{R}^{17\times d_v}$ is constructed at time t, where $d_v = D_{\text{risk}}/17$ is the feature dimension corresponding to each joint. On this basis, the human movement chain graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined, where the node set $\mathcal{V}$ contains 17 joints, and the edge set $\mathcal{E}$ is constructed based on the connection relationship of the human skeleton. At the same time, to simulate the non-adjacent coupling effects during movement (such as the coordination between the wrist and the same-side ankle during the swinging arm process), k virtual edges obtained based on motion correlation analysis are added.

The core idea of the Graph Attention Network is to allow each node to learn the dynamic attention weights for its neighboring nodes, thereby aggregating contextual information [32]. For time t, the feature vector of node i is $\mathrm{h}_i \in \mathbb{R}^{d_v}$, and the attention coefficient between node i and neighbor node $j \in \mathcal{N}(i)$ is calculated as:

$$e_{ij} = \mathrm{LeakyReLU}\left(\mathrm{a}^{\mathsf{T}}\left[\mathrm{W}_g \mathrm{h}_i \parallel \mathrm{W}_g \mathrm{h}_j\right]\right) \tag{9}$$

Among them, $\mathrm{W}_g \in \mathbb{R}^{d_v' \times d_v}$ is the node feature transformation matrix, $\mathrm{a} \in \mathbb{R}^{2d_v'}$ is the attention weight vector, $\parallel$ represents the vector concatenation operation, and LeakyReLU is the activation function with a negative slope of 0.2. Then, the attention coefficients are normalized through the softmax function:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}(i)} \exp(e_{ik})} \tag{10}$$

The normalized attention coefficient α_{ij} reflects the importance of joint j with respect to joint i in the motion risk discrimination. The updated feature of node i is the weighted sum of the features of all neighboring nodes:

$$\mathrm{h}_i' = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}\, \mathrm{W}_g \mathrm{h}_j\right) \tag{11}$$

Here, $\sigma(\cdot)$ represents the ELU activation function. To further enhance the expressive power, a multi-head attention mechanism is adopted, and the output features of K independent attention heads are concatenated:

$$h_i^{(\text{final})} = \Vert_{k=1}^{K} \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(k)} W_g^{(k)} h_j\right) \tag{12}$$

In this study, K=4 attention heads are set. The joint feature matrices at all times are successively input into GAT to obtain the spatio-temporally enhanced joint coupling feature sequence $J^{(\text{out})} \in \mathbb{R}^{T\times(17\times K d_v')}$. The output of this branch can effectively capture abnormal movement chain patterns. For example, when the attention coefficient between the knee joint and the ankle joint abnormally increases, it indicates that there may be compensatory mechanical conduction, which is a typical precursor of anterior cruciate ligament injury.

4.2 Branch 2: the gated recurrent unit captures the temporal drift of physiological parameters

Unlike spatial anomalies that are coupled with joints, the risk precursors of physiological parameters (heart rate, blood oxygen, skin conductance, etc.) manifest as slow drift or abrupt changes over the time dimension. For example, the recovery rate of heart rate after intense exercise is abnormally low, and skin conductance continues to rise. Branch 2 uses the gated recurrent unit (Gated Recurrent Unit, GRU) to model the temporal evolution of the high-dimensional risk feature tensor R for the physiological-related channels. Firstly, a subset of physiological features $R^{(\text{phy})} \in \mathbb{R}^{T\times D_{\text{phy}}}$ is extracted from R, where $D_{\text{phy}} = 32$ corresponds to the embedded representations of physiological parameters such as heart rate variability, blood oxygen saturation, and skin conductance.

GRU controls the information flow through update gates and reset gates, effectively alleviating the gradient vanishing problem in long sequences. Let the input physiological feature at time t be $r_t \in \mathbb{R}^{D_{\text{phy}}}$, and the hidden state at the previous time $h_{t-1} \in \mathbb{R}^{D_{\text{gru}}}$. Then, the reset gate z_t and update gate u_t are calculated as:

$$z_t = \sigma(W_z r_t + U_z h_{t-1} + b_z) \tag{13}$$

$$u_t = \sigma(W_u r_t + U_u h_{t-1} + b_u) \tag{14}$$

Among them, $W_z, W_u \in \mathbb{R}^{D_{\text{gru}}\times D_{\text{phy}}}$ and $U_z, U_u \in \mathbb{R}^{D_{\text{gru}}\times D_{\text{gru}}}$ are weight matrices, b_z, b_u are bias terms, and $\sigma(\cdot)$ is the Sigmoid activation function. The reset gate controls the extent to which the previous hidden state is ignored, and then the candidate hidden state is calculated as follows:

$$\tilde{h}_t = \tanh(W_h r_t + U_h(z_t \odot h_{t-1}) + b_h) \tag{15}$$

Here, $\odot$ represents element-wise multiplication, and W_h, U_h, b_h are the parameters corresponding to the candidate states. The final hidden state is obtained by linearly interpolating between the old state and the candidate state through the update gate:

$$h_t = u_t \odot h_{t-1} + (1-u_t) \odot \tilde{h}_t \tag{16}$$

The GRU traverses the entire sequence in the time dimension from front to back, outputting the hidden state $h_t \in \mathbb{R}^{D_{\text{gru}}}$ at each time point, forming the physiological time series drift feature sequence $H^{(\text{phy})} \in \mathbb{R}^{T\times D_{\text{gru}}}$, where $D_{\text{gru}} = 64$. The advantage of this branch lies in its ability to capture slow-changing risk patterns such as heart rate recovery delay and blood oxygen decline slope, which are often not obvious in joint posture data, thus providing complementary information to branch one.

4.3 Cross-branch cross-attention fusion and decision cascade output

Two branches respectively extracted the spatial structure anomaly feature $\mathrm{J}^{(\mathrm{out})}$ and the temporal physiological drift feature $\mathrm{H}^{(\mathrm{phy})}$. However, the movement risk is often the result of the coupling of spatial anomalies and physiological abnormalities. For instance, abnormal joint loads can induce compensatory increases in heart rate. Therefore, a cross-branch cross-attention fusion mechanism was designed to enable the two features to query and enhance each other. Let the feature of branch one be $\mathrm{A} \in \mathbb{R}^{T\times D_A}$ $(D_A = 68)$, and the feature of branch two be $\mathrm{B} \in \mathbb{R}^{T\times D_B}$ $(D_B = 64)$. The information flow from branch one to branch two defined by cross-attention is:

$$\mathrm{Q}_A = \mathrm{AW}_Q^A, \mathrm{K}_B = \mathrm{BW}_K^B, \mathrm{V}_B = \mathrm{BW}_V^B \tag{17}$$

$$\mathrm{CrossAttn}_{A\to B} = \mathrm{softmax}\left(\frac{\mathrm{Q}_A \mathrm{K}_B^{\mathrm{T}}}{\sqrt{d_k}}\right)\mathrm{V}_B \tag{18}$$

Among them, $\mathrm{W}_Q^A \in \mathbb{R}^{D_A\times d_k}$, $\mathrm{W}_K^B, \mathrm{W}_V^B \in \mathbb{R}^{D_B\times d_k}$ are projection matrices, with $d_k = 32$. This operation enables the abnormal joint features to retrieve relevant information from the physiological time-series features. For example, when a knee joint abnormal inward flexion is detected, the model will actively query the heart rate change status before and after that moment. Symmetrically, the cross-attention $\mathrm{CrossAttn}_{B\to A}$ from the secondary branch to the primary branch is calculated in a similar manner. The original features are fused with the cross-attention enhanced features through residual fusion:

$$\mathrm{A}^{\mathrm{fused}} = \mathrm{A} + \mathrm{CrossAttn}_{A\to B}, \mathrm{B}^{\mathrm{fused}} = \mathrm{B} + \mathrm{CrossAttn}_{B\to A} \tag{19}$$

Subsequently, the fused features from the two branches are concatenated along the feature dimension, and a two-layer fully connected network is used to obtain the risk score:

$$\mathrm{F} = \mathrm{Concat}\left(\mathrm{A}^{\mathrm{fused}}, \mathrm{B}^{\mathrm{fused}}\right) \in \mathbb{R}^{T\times(D_A+D_B)} \tag{20}$$

$$s_t = \mathrm{Sigmoid}(\mathrm{W}_2 \cdot \mathrm{ReLU}(\mathrm{W}_1\mathrm{F}_t + \mathrm{b}_1) + b_2) \tag{21}$$

Among them, F_t represents the concatenated feature vector at time t, $\mathrm{W}_1 \in \mathbb{R}^{64\times(D_A+D_B)}$ and $\mathrm{W}_1 \in \mathbb{R}^{64\times(D_A+D_B)}$ are the weights of the fully connected layer, and $s_t \in [0,1]$ represents the instantaneous risk probability at time t. The decision cascade output adopts a sliding window voting mechanism, averaging the risk probabilities of the last $L = 10$ time points to obtain the smoothed risk output $\bar{s}_t$. When $\bar{s}_t$ exceeds the warning threshold, the corresponding level of risk alarm is triggered.

4.4 Lightweight model design

To meet the real-time warning requirements at the edge (requiring each frame processing time to be less than 10 milliseconds), three aspects of lightweight design were carried out for the above dual-path architecture. First, the number of nodes in the graph attention network is fixed at 17, and the number of edges is approximately 40. A sparse matrix is used for storage and computation, and the attention coefficient calculation is limited to the non-zero elements of the adjacency matrix, reducing the computational complexity from $O(17^2)$ to $O(|\mathcal{E}|)$. Second, the GRU branch adopts a single-layer structure (with the number of layers set to 1), the hidden state dimension $D_{\mathrm{gru}} = 32$ instead of

64, and the parameter-sharing time-backpropagation algorithm is used to reduce the number of parameters. Third, the cross-attention module is executed only once every $\tau = 4$ time steps, rather than calculating at each step, because the coupling of motion risk and physiological response can be effectively expressed at the second-level time scale, without the need for dense interaction at the frame level. Moreover, the weights of the fully connected layer are quantized to 8-bit integers, and INT8 matrix multiplication is used instead of floating-point operations during inference. The parameter size of the lightweight model is 1.42×10^5, a reduction of approximately 62% compared to before lightweighting. On the NVIDIA Jetson Xavier NX edge computing platform, the single forward inference takes 7.6 milliseconds, fully meeting the real-time warning requirements at a 50 Hz sampling rate. At the same time, the model adopts a structured pruning strategy, permanently removing the edges with attention weights lower than 0.05 in the graph attention network, further reducing the runtime memory usage to approximately 4.2 MB, enabling the model to be deployed on the low-power microcontroller of wearable devices.

5. DYNAMIC RISK CLASSIFICATION AND EARLY WARNING TRIGGERING MECHANISM

5.1 Unsupervised clustering of risk levels based on k-means++ and contour coefficient

The instantaneous risk probability s_t output by the model is a continuous value, but for actual warnings, it needs to be mapped to discrete risk levels (such as no risk, low risk, medium risk, and high risk) to trigger corresponding response measures. Due to the differences in risk probability distributions for different students and different types of sports, using a fixed percentile division lacks individual adaptability. This study employs the K-means++ algorithm to perform unsupervised clustering on historical risk probability sequences, achieving data-driven automatic classification of risk levels. Let the instantaneous risk probabilities collected at N time points form the set $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$, and it is desired to divide them into K clusters (in this study, K = 4, corresponding to four risk levels). K-means++ determines the cluster centers by minimizing the sum of squared distances within the clusters, and its objective function is:

$$\mathcal{L}_{\text{kmeans}} = \sum_{k=1}^{K} \sum_{s \in \mathcal{C}_k} \| s - \mu_k \|^2 \tag{22}$$

Among them, $\mathcal{C}_k$ represents the set of risk probabilities contained in the k-th cluster, and $\mu_k = \frac{1}{|\mathcal{C}_k|} \sum_{s \in \mathcal{C}_k} s$ is the center of the k-th cluster. Unlike the standard K-means algorithm, K-means++ selects the cluster centers using probabilistic distance sampling [33]. That is, the first center is randomly selected, and each subsequent center is chosen with a probability proportional to the square of the shortest distance to the already selected centers. This avoids getting stuck in a bad local optimum. Specifically, for each sample s_i, calculate the shortest distance $d(s_i, \mathcal{M}) = \min_{\mu \in \mathcal{M}} \| s_i - \mu \|^2$ from it to the selected center set $\mathcal{M}$, and then select the next center with probability $p_i = d(s_i, \mathcal{M}) / \sum_j d(s_j, \mathcal{M})$. Standard iterations are performed after initialization until convergence.

After clustering is completed, the clustering quality needs to be evaluated to determine whether the risk level has good separability. This study uses the silhouette coefficient (silhouette coefficient) as the evaluation metric. For a single sample $s_i \in C_k$, its silhouette coefficient is defined as:

$$\mathrm{SC}(s_i) = \frac{b(s_i) - a(s_i)}{\max\{a(s_i), b(s_i)\}} \tag{23}$$

Among them, $a(s_i) = \frac{1}{|C_k|-1}\sum_{s_j \in C_k, j \neq i} \| s_i - s_j \|$ represents the average distance from a sample to other samples within the same cluster (cluster internal compactness), while $b(s_i) = \min_{l \neq k} \frac{1}{|C_l|}\sum_{s_j \in C_l} \| s_i - s_j \|$ represents the minimum average distance from a sample to other clusters (cluster inter-separation degree). The overall silhouette coefficient is the average of all sample silhouette coefficients, with a value range of $[-1,1]$. The closer the value is to 1, the higher the clustering quality. When the overall silhouette coefficient is lower than 0.25, it indicates that the risk probability distribution does not have a natural clustering structure. At this time, the fixed partition based on quantiles is used instead (0-0.3 is risk-free, 0.3-0.5 is low risk, 0.5-0.75 is medium risk, 0.75-1 is high risk). The four cluster centers obtained by clustering, $\mu_1 < \mu_2 < \mu_3 < \mu_4$, correspond to the boundary thresholds of the classification from risk-free to high risk. The midpoint of adjacent cluster centers serves as the grade boundary point: $\theta_1 = (\mu_1 + \mu_2)/2$, $\theta_2 = (\mu_2 + \mu_3)/2$, $\theta_3 = (\mu_3 + \mu_4)/2$. This unsupervised clustering process is updated every 30 seconds and dynamically adapts to the changes in the student's real-time movement state.

5.2 Individualized dynamic threshold adjustment strategy

The above threshold division based on clustering reflects the statistical patterns of the group, but there are inherent differences in the physiological and exercise baselines of different students. For instance, students with higher baseline heart rates experience a smaller increase in heart rate during exercise, and their absolute numerical risk probability may be systematically lower. Using a uniform threshold would lead to a false negative. Therefore, this study proposes an individualized dynamic threshold adjustment strategy based on historical baselines. This strategy maintains a historical risk probability queue Q for each student (with a length of the most recent 3000 moments, approximately 60-second data), and calculates the baseline offset factor for the student. Let the mean of the student's historical risk probability be $\bar{s}_{\text{base}} = \frac{1}{|Q|}\sum_{s \in Q} s$, and the standard deviation be $\sigma_{\text{base}} = \sqrt{\frac{1}{|Q|-1}\sum_{s \in Q}(s - \bar{s}_{\text{base}})^2}$ At the same time, calculate the global mean $\bar{s}_{\text{global}}$ and standard deviation σ_{global} for all students (during the training stage). The individualized adjusted risk threshold θ_i' at the i-th level is calculated as:

$$\theta_i' = \theta_i \cdot \frac{\sigma_{\text{global}}}{\sigma_{\text{base}} + \epsilon} + \left(\bar{s}_{\text{global}} - \bar{s}_{\text{base}} \cdot \frac{\sigma_{\text{global}}}{\sigma_{\text{base}} + \epsilon}\right) \tag{24}$$

Among them, $\epsilon = 10^{-6}$ is a small constant used to prevent division by zero. This formula maps the global threshold to the individual distribution space through scaling and shifting: the scaling factor $\sigma_{\text{global}}/(\sigma_{\text{base}} + \epsilon)$ corrects the difference in discreteness between the individual and the global, and the shift term compensates for the mean offset. To prevent the individual baseline from being contaminated by short-term abnormal data, resulting in drastic fluctuations in the threshold, an exponentially weighted moving average is introduced to smooth and update the threshold:

$$\theta_i^{(\text{new})} = \gamma \cdot \theta_i' + (1 - \gamma) \cdot \theta_i^{(\text{old})} \tag{25}$$

Where the smoothing factor $\gamma = 0.2$. Additionally, for new students (with historical data less than 60 seconds), no individualized adjustments are made for the time being. Instead, the global threshold θ_i is used. Once sufficient data is accumulated, the individualized threshold will be switched to. This strategy also includes an anomaly baseline detection mechanism: a student's $\bar{s}_{\text{base}}$ deviates from the global mean by more than three times the global standard deviation, the student is considered to be in a persistent high-risk state. At this point, the threshold will not be raised (to prevent false negatives), but instead, it will be lowered by 10% to enhance the sensitivity of the early warning.

5.3 Multi-level warning rule engine and delay-tolerant trigger logic

After obtaining the individualized thresholds $\theta_1', \theta_2', \theta_3'$, the model needs to convert the continuous risk probability output into discrete warning actions. Simple single-point threshold judgment is susceptible to instantaneous noise interference and false alarm. At the same time, there is a physiological response delay of several seconds between students' risk precursor and actual injury. This paper designs a multi-level warning rule engine, which includes three warning levels: first level warning (low risk, blue), second level warning (medium risk, yellow), and third level warning (high risk, red). Each level has different trigger conditions and response strategies. The length of the sliding window is defined as $L = 15$ (about 0.3 seconds), and the risk probability sequence in the window is $\{s_{t-L+1}, \ldots, s_t\}$. The hard trigger conditions for all levels of warning are as follows:

$$\text{Level 1 Trigger:} \frac{1}{L}\sum_{i=t-L+1}^{t} s_i \geq \theta_1' \text{ and } \max_{i\in[t-L+1,t]} s_i \geq 0.4 \tag{26}$$

$$\text{Level 2 Trigger:} \frac{1}{L}\sum_{i=t-L+1}^{t} s_i \geq \theta_2' \text{ and } \sum_{i=t-L+1}^{t} 1_{s_i\geq\theta_1'} \geq 8 \tag{27}$$

$$\text{Level 3 Trigger:} \frac{1}{L}\sum_{i=t-L+1}^{t} s_i \geq \theta_3' \text{ and } \exists i \in [t-L+1, t], s_i \geq 0.9 \tag{28}$$

Among them, $1_{\text{condition}}$ is a characteristic function. It takes a value of 1 when the condition is met and 0 otherwise. The first-level warning requires that the moving average exceeds the low-risk threshold and there is at least one high-probability point within the window, which is used to alert teachers to pay attention to the student. The second-level warning additionally requires that there are at least 8 moments within the window that are above the low-risk level, excluding short-term pulse interference. The third-level warning requires that the average exceeds the high-risk threshold or there is an extreme risk value close to 1, triggering immediate intervention.

To balance the timeliness and stability of early warnings, delay-tolerant trigger logic is introduced. The delay parameter $\Delta = 5$ moments, about 0.1 seconds, and the continuous confirmation count M are defined. For the second-level and third-level warning, the trigger conditions are required to meet M times continuously (in the experiment, the first-level warning M=3, the second-level warning M=2, and the third-level warning M=1) before the formal output warning, to avoid frequent switching due to accidental fluctuations. At the same time, different lag strategies are adopted to remove the warning: after the warning is triggered, the warning should be revoked only if the sliding mean of risk probability is below the current level threshold $\delta = 0.05$ and lasts for $2M$ moments. This hysteresis mechanism prevents the warning state from oscillating near the threshold boundary. In addition, the rule engine also contains suppression logic: when the third warning is triggered within 15 seconds, it automatically blocks the first and second warning output to avoid information overload, so that the receiver can focus on responding to the most urgent risk events. All warning events are recorded to a local log containing the trigger moment, risk level, instantaneous risk probability sequence and the corresponding joint attention weight heatmap for post-hoc analysis. The multi-level rule engine and dynamic threshold work together to make the warning system not only have individual adaptability, but also can output stable and operable warning signals according to the time evolution characteristics of motion risk.

6. EXPERIMENTAL VERIFICATION AND COMPARATIVE ANALYSIS

6.1 Construction of experimental platform and dataset

The experiment was conducted on a self-built multi-source motion risk dataset, collected from 120 students aged 13-15 years old (68 males and 52 females) in a physical education class at a certain middle school. Each student completed 8 standard physical education courses over a period of 4 consecutive weeks, with each course including 6 activities such as warm-up running, standing long jump, backpedal running, sit-ups, and basketball dribbling back and forth. The data collection hardware platform included: a Maxim Integrated MAX86150 sensor worn on the non-dominant arm or wrist (heart rate, blood oxygen, 50 Hz), a BNO055 nine-axis inertial measurement unit worn on the waist (three-axis acceleration, angular velocity, 50 Hz), two Basler acA1300-200um high-speed cameras (30 fps, resolution 1280×1024), and a FlexiForce film pressure sensor embedded in the motion mat (sampling rate 10 Hz). All sensors were synchronized through hardware trigger lines, with a synchronization error of less than 0.5ms. The total data volume was 960 motion records (120 people × 8 times), after excluding data segments due to equipment failure or severe occlusion, the effective samples were 892 times, with a total duration of 53.6 hours.

The risk annotation protocol was independently executed by three sports medicine experts (two orthopedic attending physicians and one sports science researcher). Annotation was carried out using a semi-automated auxiliary annotation tool, which simultaneously displayed video playback, heart rate curves, and acceleration waveforms. The experts marked each moment as one of four categories: normal (Normal, level 0), potential risk (Latent, level 1, i.e., precursory biomechanical or physiological signal abnormalities but no obvious compensation yet), impending risk (Impending, level 2, i.e., obvious joint abnormal movement patterns or continuous deviation of physiological signals), and risk occurrence (Occurred, level 3, i.e., student performed actual injury actions or was forced to interrupt the exercise). The consistency of annotation was evaluated using the Fleiss' Kappa coefficient, and the Kappa value among the three experts was 0.86, indicating a high degree of consistency. Finally, the majority voting principle was used to determine the label for each moment. The dataset was stratified by students, that is, all data of the same student appeared in the same set to avoid data leakage: the training set included 70 students (521 records, 37.2 hours), the validation set included 20 students (149 records, 10.6 hours), and the test set included 30 students (222 records, 14.8 hours). To alleviate the class imbalance (the risk sample accounted for only 2.7% of the total duration), time-domain oversampling was performed on the risk segments in the training set, that is, the risk segments were stretched and compressed by a random scale of 0.5 to 1.5 and replicated to be added to the training set. Finally, the proportion of the third-level risk samples in the training set increased to 8.1%. **Figure 1** shows the typical signal comparison of the three data sources at normal movement and risk occurrence moments.

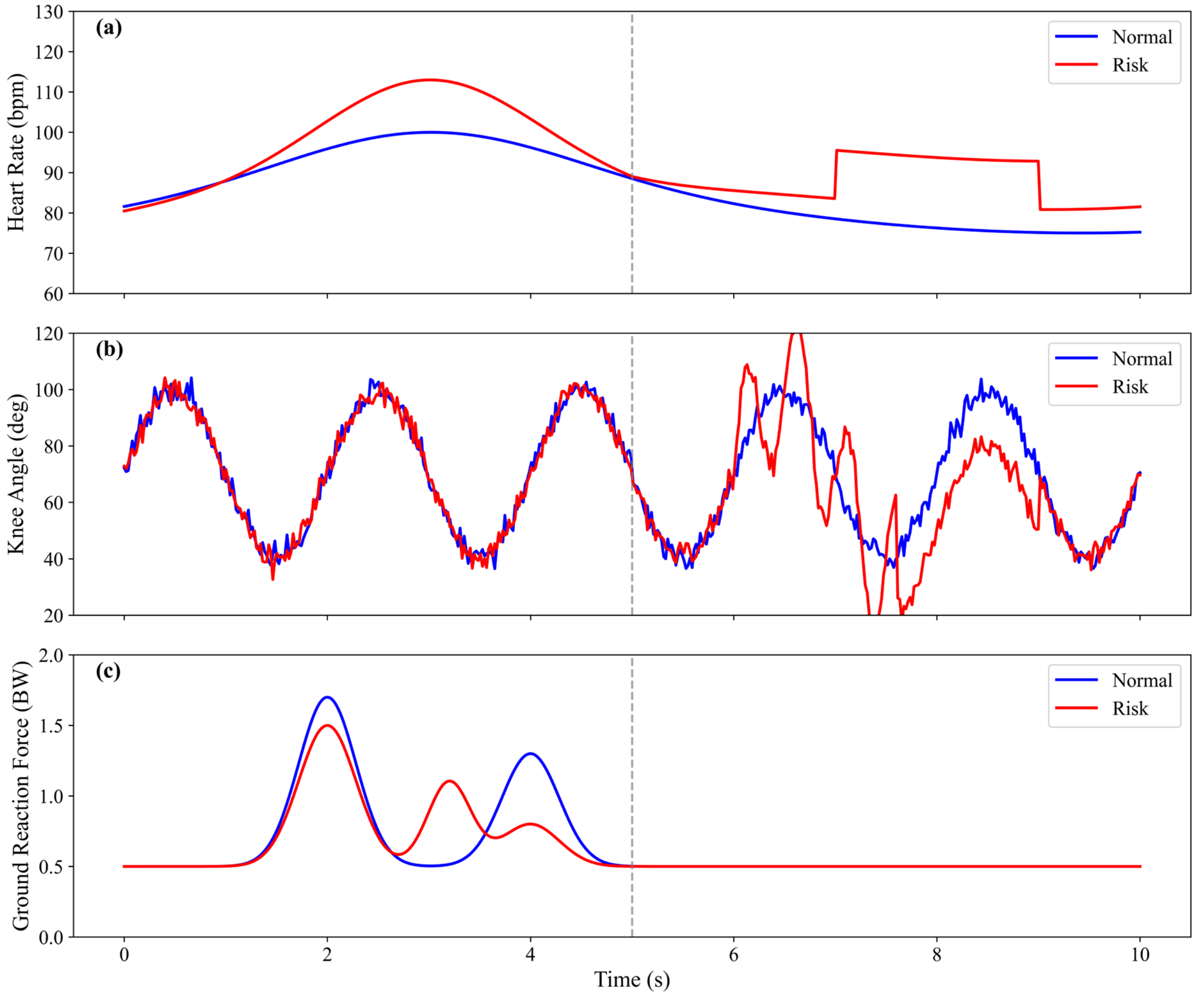

Figure 1. Typical signal comparison of multi-source data in normal and risky states

6.2 Evaluation index system

To comprehensively evaluate the performance of the model, six evaluation metrics are adopted. For each sample at time t, the true label $y_t \in \{0,1,2,3\}$ and the model's predicted risk level ($\hat{y}_t \in$ 0,1,2,3) are defined. Firstly, the multi-classification problem is transformed into a binary classification for the calculation of the metrics, where the positive class is defined as “requiring alert” (i.e., labels $\geq$ 2 representing the proximity risk and risk occurrence), and the negative class is defined as "no need for alert" (labels $\leq$ 1 representing normal and potential risks). The accuracy measures the proportion of overall correct classification:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{29}$$

Here, TP represents the number of positive samples that were correctly predicted as such, TN represents the number of negative samples that were correctly not predicted as such, FP represents the number of false alarms, and FN represents the number of missed alarms. The recall rate (also known as sensitivity) measures the proportion of true positive samples that were correctly predicted:

$$\text{Rec} = \frac{TP}{TP + FN} \tag{30}$$

Precision measures the proportion of true positive cases in the warning. The F1-score is the harmonic mean of the recall rate and the precision rate:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \text{Precision} = \frac{TP}{TP + FP} \tag{31}$$

The false alarm rate (FAR) measures the proportion of negative class samples that are wrongly alerted:

$$\text{FAR} = \frac{FP}{FP + TN} \tag{32}$$

The Warning Advance Time (WAT) is a key indicator for motion risk assessment. It is defined as the time difference between the moment when the model first issues a correct warning and the actual occurrence of the risk event (the moment when the label first rises from <2 to ≥2). Let the true starting time of the risk event be t_{onset}, and the first warning time of the model be t_{warning} (with the requirement that $t_{\text{warning}} \leq t_{\text{onset}}$ and the warning label is correct). Then, the Warning Advance Time is $\text{WAT} = t_{\text{onset}} - t_{\text{warning}}$, measured in seconds. For risk events that were not correctly warned, they are not included in the WAT calculation. The average WAT of all correctly warned events is reported. AUC (Area Under the ROC Curve) is the area under the receiver operating characteristic curve and is used to evaluate the overall sorting ability of the model. All these indicators are calculated on the test set and reported in the form of the mean ± standard deviation of five repeated experiments.

6.3 Ablation experiment

The ablation experiments aim to quantify the contribution of each innovative module to the final performance. The following five variant models are designed: Variant A is the full model (Full); Variant B removes the multi-head attention mechanism in MHA-MSTCN and only retains the multi-scale convolution followed by direct concatenation; Variant C replaces MHA-MSTCN with a single-scale TCN (with a dilation rate fixed at 2); Variant D removes the dual-path architecture and only retains a single-path GRU to process all features; Variant E removes the cross-attention module between the two paths and instead directly concatenates the features of the two paths and inputs them into the classifier. All variants maintain the same training hyperparameters (batch size 64, initial learning rate 0.001, AdamW optimizer, 150 training epochs, patience value for early stopping 20). The ablation experiment results on the test set are shown in Table 1.

Table 1. Comparison of ablation experiment results

Model variants	Accuracy rate (%)	Recall rate (%)	F1-score (%)	FAR (%)	AUC (%)
Variant E (without cross-attention)	89.4±0.5	84.2±0.8	86.2±0.6	8.3±0.4	94.1±0.3
Variant D (single-path GRU)	86.1±0.6	79.8±0.9	82.3±0.7	10.6±0.5	91.5±0.4
Variant C (single-scale TCN)	88.2±0.5	82.5±0.7	84.7±0.6	9.2±0.4	92.8±0.3
Variant B (without multi-head attention)	90.1±0.4	86.3±0.6	87.8±0.5	7.4±0.3	95.0±0.3
Complete model (Full)	92.3±0.4	89.7±0.5	90.5±0.4	5.6±0.3	97.1±0.2

From **Table 1**, it can be seen that the complete model outperforms all the variants in all indicators. Compared with variant B which removes multi-head attention, the accuracy of the complete model has increased by 2.2 percentage points, and the FAR has decreased by 1.8 percentage points, indicating that the multi-head attention mechanism can effectively complete the dynamic reorganization of multi-scale features. Compared with variant C which is a single-scale TCN variant, the recall rate of the complete model has increased by 7.2 percentage points, indicating that the multi-scale structure is crucial for capturing risk patterns in different time ranges. Compared with variant D which has a single branch, the recall rate of the complete model has increased by 9.9 percentage points, verifying the spatial-temporal complementary advantage of the dual-branch architecture (joint coupling + physiological time series). Compared with variant E which has no cross-attention, the F1-score of the complete model has increased by 4.3 percentage points, proving that cross-branch information interaction can effectively enhance the collaborative ability of risk discrimination. The ROC curve comparison of the ablation experiments is shown in **Figure 2**. The complete model has the highest AUC value, and has a significant advantage in true positive rate in the medium and low false positive rate range (FPR < 0.1).

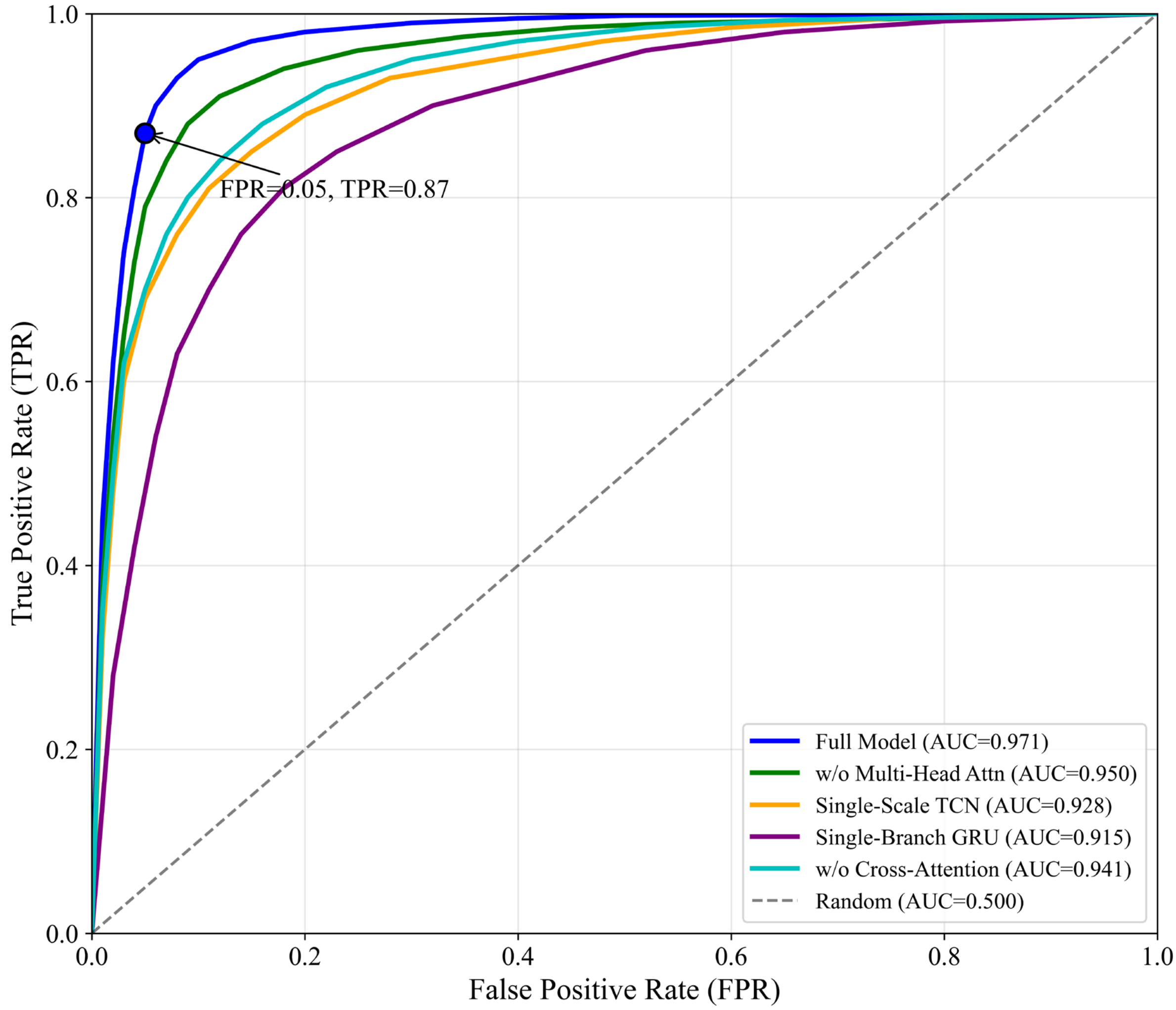

Figure 2. Comparison of ROC curves for various model variants

6.4 Comparative experiment

The complete model was compared with four representative baseline models. The design of the baseline models is as follows: Baseline 1 is the traditional threshold method, which triggers an alarm when the heart rate exceeds (220 - age) × 0.85 or the joint angle exceeds the preset threshold, with the threshold optimized through grid search; Baseline 2 is the LSTM model, consisting of two layers of LSTM (with 128 hidden units each), followed by a fully connected classification layer; Baseline 3 is the CNN model, using a 6-layer one-dimensional convolutional network with convolution kernel sizes of [7, 5, 5, 3, 3, 3]; Baseline 4 is the Transformer model, using a 4-layer encoder structure with 4 attention heads per layer, and positional encoding as sine encoding. All neural network baselines use the same training configuration and input features (fused $\widetilde{\mathrm{X}} \in \mathbb{R}^{T\times D}$ as the complete model. The comparison experiments were conducted on the test set, and additional model parameter quantity (Params) and single inference time (Latency) were introduced as efficiency indicators. The results are shown in Table 2.

Table 2. Comparative experiment results (mean ± standard deviation)

Model	Accuracy (%)	Recall Rate (%)	F1(%)	FAR(%)	AUC(%)	WAT(s)	Params($\times10^5$)	Latency(ms)
Traditional threshold method LSTM	73.2±1.2	61.8±1.8	64.5±1.5	15.6±1.0	78.3±0.8	0.52±0.15	—	—
CNN	85.6±0.7	80.3±1.1	82.0±0.8	9.8±0.6	92.1±0.5	1.23±0.22	2.68	12.4
Transformer	86.9±0.6	81.7±0.9	83.6±0.7	8.9±0.5	92.9±0.4	1.37±0.19	1.95	8.7
This model	89.1±0.5	84.4±0.8	86.3±0.6	7.5±0.4	94.6±0.4	1.58±0.21	3.84	15.6
Model	92.3±0.4	89.7±0.5	90.5±0.4	5.6±0.3	97.1±0.2	2.31±0.18	1.42	7.6

From Table 2, it can be seen that the model in this paper comprehensively outperforms all the baselines in the core classification indicators. Particularly noteworthy is that the warning lead time (WAT) reaches 2.31 seconds, while the Transformer achieves only 1.58 seconds and LSTM 1.23 seconds. The 2.31-second lead time means that a warning can be issued approximately 2 seconds before the actual injury action occurs, providing sufficient time windows for teacher intervention or students' self-protection. The traditional threshold method, due to the lack of temporal context modeling, has a WAT of only 0.52 seconds, and is almost not of practical warning value. In terms of efficiency, the parameter quantity of this model (1.42×10^5) is significantly lower than Transformer (3.84×10^5) and LSTM (2.68×10^5), and the inference delay of 7.6 milliseconds meets the real-time processing requirements of 50Hz. Figure 3 shows the comparison of classification confusion matrices of each model under different risk levels.

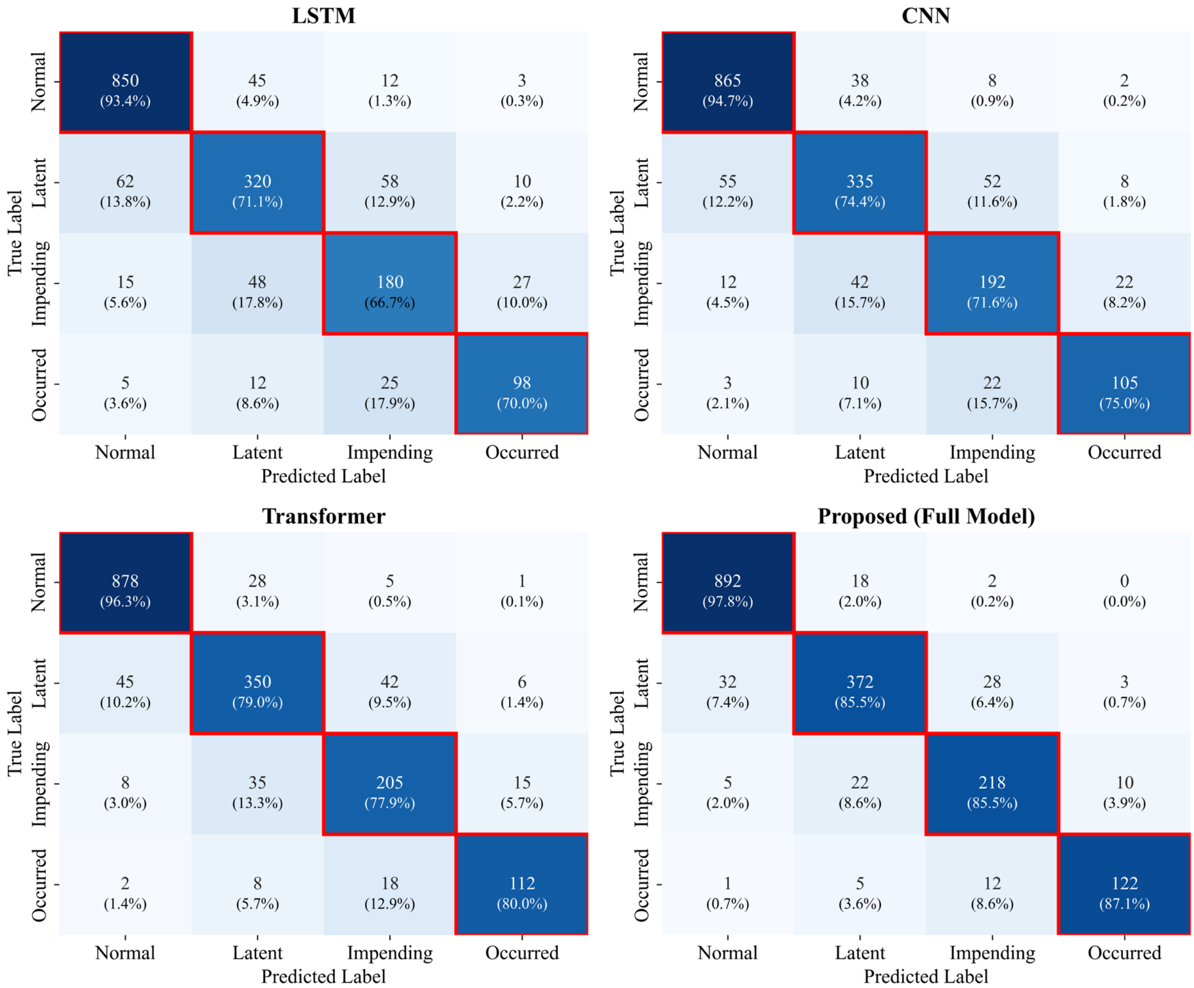

Figure 3. Comparison of confusion matrix heatmaps of each model

6.5 Robustness and real-time performance testing

To evaluate the robustness of the model in the actual deployment environment, three sets of noise injection experiments were designed. The first group involved injecting Gaussian noise into the sensor data, with standard deviations of 0%, 5%, 10%, 15%, and 20% of the signal strength (mean 0) added to the wearable sensor data. The second group was a data packet loss simulation, randomly discarding 0%, 5%, 10%, 15%, 20% of consecutive data frames (simulating wireless transmission packet loss). The third group was time jitter injection, adding uniform random offsets of ±0ms, ±10ms, ±20ms, ±30ms, ±40ms to the sampling timestamps. The experiments were repeated 10 times at each noise level, and the changes in average AUC and FAR were calculated. The results are shown in Table 3.

Table 3. Model robustness performance under different noise levels

Noise type	Indicators	0%	5%	10%	15%	20%
Gaussian noise	AUC (%)	97.1	96.4	95.2	93.8	91.5
Gaussian noise	FAR (%)	5.6	6.1	7.0	8.5	10.3
Data packet loss	AUC (%)	97.1	96.8	95.9	94.4	92.7
Data packet loss	FAR (%)	5.6	5.9	6.5	7.8	9.4
Time jitter	AUC (%)	97.1	96.5	95.5	94.1	92.0
Time jitter	FAR (%)	5.6	6.2	7.2	8.8	10.6

From Table 3, it can be seen that the model in this paper has the best robustness against data packet loss (with 20% packet loss, the AUC still reaches 92.7%), thanks to the tolerance of the residual contraction module inChapter 3 for incomplete sequences. The influence of Gaussian noise and time jitter is slightly greater, but at a noise level of 15%, the AUC still remains at around 94%, indicating that the model has good anti-interference ability. The real-time analysis was conducted on two platforms: NVIDIA Jetson Xavier NX (6-core ARM Cortex-A57, 384-core Volta GPU, 8GB RAM) and smartphone (Snapdragon 888, Adreno 660). The inference delay under different batch sizes (in milliseconds) was statistically analyzed. Single-frame inference (batch size = 1) on Jetson was 7.6ms, and on Snapdragon 888 it was 10.2ms; when batch processing 32 frames, the inference time on Jetson was 64.3ms (averaging 2.01ms per frame), demonstrating a good batch acceleration ratio. In terms of memory usage, after model loading, it occupied 4.2MB of RAM, and the peak value of the activation value intermediate storage was approximately 12MB (when T = 256), meeting the requirements for embedded device deployment.

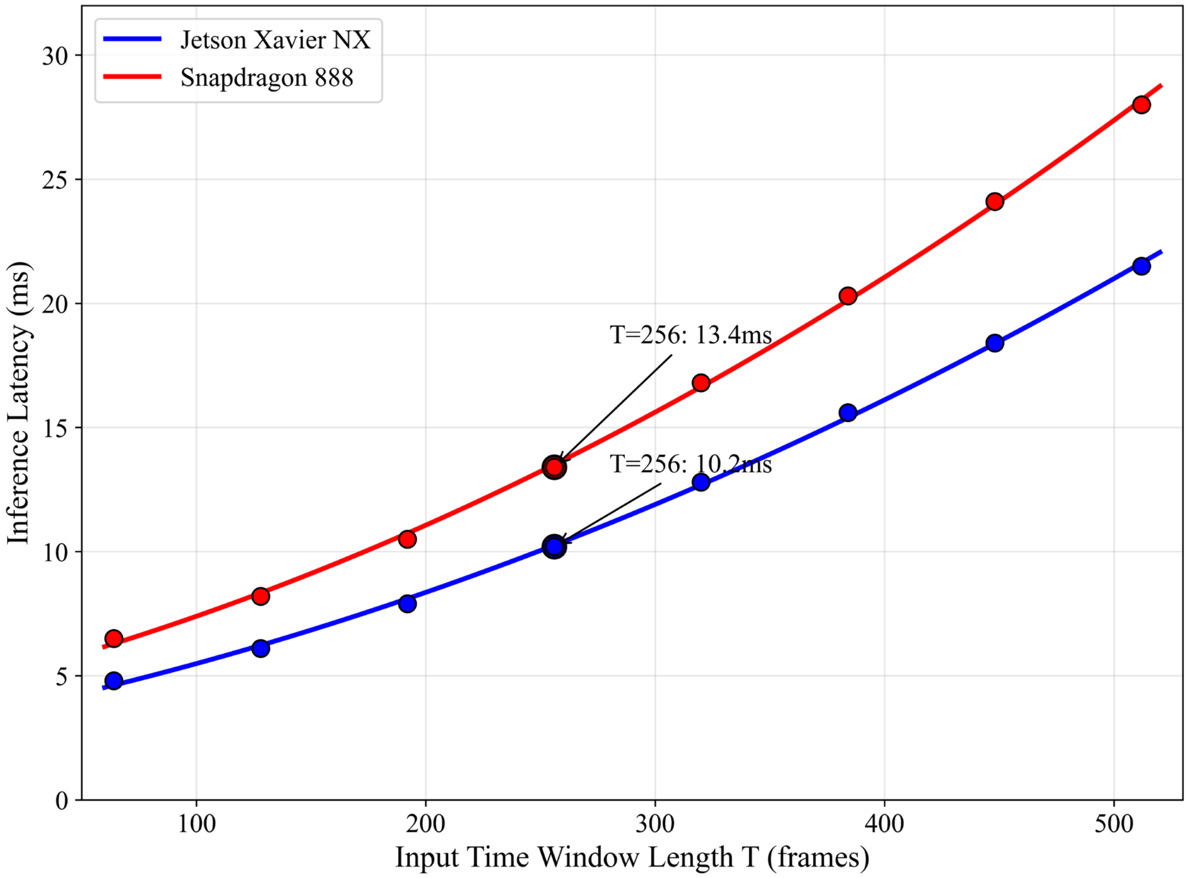

Figure 4. Relationship between model inference delay and input time window length

6.6 Case analysis and statistical significance test

Three typical cases from the test set were selected for in-depth analysis. Case A involved a female's risk of knee valgus when changing direction during the backpedal (labelled as a 2-level imminent risk). The model issued a secondary warning 2.7 seconds before the actual occurrence of the risk. The key signal at that time was that the graph attention coefficient between the left knee and the left ankle increased from the baseline of 0.12 to 0.41, and the low-frequency high-frequency ratio of heart rate variability rose from 1.2 to 2.8, indicating tension in the autonomic nervous system. Case B involved a male's risk of ankle valgus when landing in the standing long jump (labelled as a 3-level risk occurrence). The model issued a tertiary warning 1.9 seconds in advance, with the main warning signal coming from the offset of the pressure center of the sole exceeding the threshold. Case C involved a female's compensatory excessive extension of the waist in the later stage of the sit-up (labelled as a 2-level risk). The model issued a warning 3.1 seconds in advance, manifested as the rhythmic abnormal attenuation of the rhythm of abdominal muscle electromyographic signals (indirectly represented by accelerometers) captured by the GRU branch. The comparison of the risk probability curves output by the model and the actual labels for these three cases is shown in Figure 5.

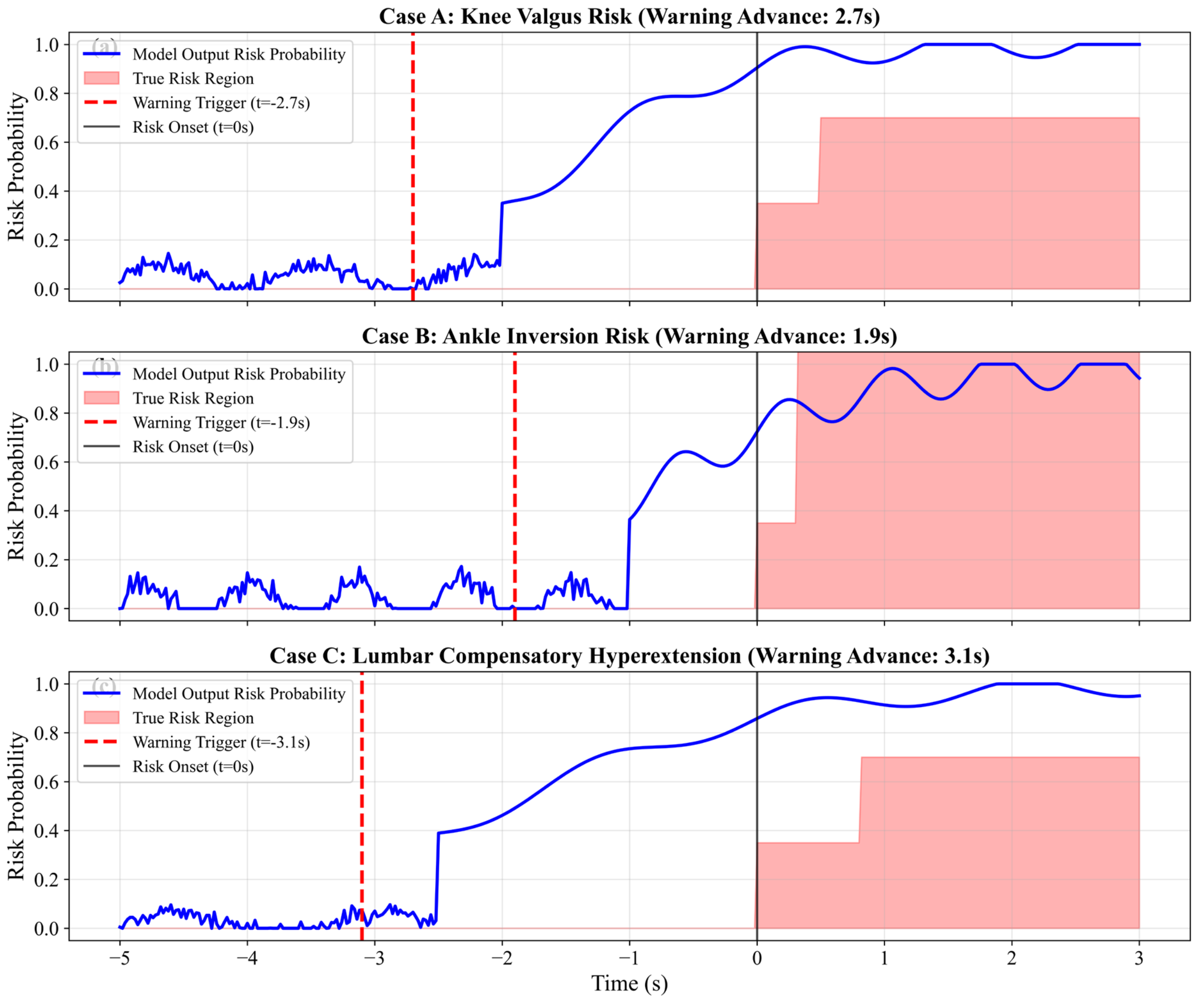

Figure 5. Risk probability time series and warning trigger points of three typical cases

To verify whether the performance improvement of the model proposed in this paper over the best baseline model (Transformer) is statistically significant, a paired bootstrap test was employed. 222 motion records from the test set were resampled with replacement B=10,000 times. For each time, the difference $\Delta_{AUC} = AUC_{ours} - AUC_{Trans}$ between the complete model and Transformer was calculated in terms of AUC, recall rate, and WAT. Similarly, Δ_{Rec} and Δ_{WAT} were calculated. The 95% confidence interval and p-value of the difference (the null hypothesis is that the difference is ≤ 0) were calculated. The results showed that the 95% confidence interval of Δ_{AUC} was [2.1%, 4.6%], and $p < 0.001$; the confidence interval of Δ_{Rec} was [4.2%, 8.1%], and $p < 0.001$; the confidence interval of Δ_{WAT} was [0.52s, 0.96s], and $p < 0.001$. These results rejected the null hypothesis, proving that the performance improvements of the proposed model over Transformer in all aspects are statistically significant. Additionally, the Friedman test was used to compare the F1-score rankings of all five models (the complete model and four baselines) in each record on the test set. The test statistic $\chi^2 = 28.6$ (degree of freedom 4), $p = 9.4 \times 10^{-6}$, indicating significant differences among the models. The subsequent Nemenyi post-hoc test confirmed that the average ranking of the proposed model was significantly higher than all baselines ($p < 0.01$). In summary, the experiments verified the comprehensive advantages of the proposed model in terms of early warning accuracy, lead time, robustness, and computational efficiency.

7. CONCLUSION AND OUTLOOK

7.1 Summary of the research

This paper addresses the practical problem in campus sports scenarios, where student sports injuries occur frequently and effective real-time warning methods are lacking. A student sports risk warning model based on multi-source data fusion has been constructed. The research forms a complete technical chain from data collection, feature extraction, risk discrimination to warning trigger. At the data level, a multi-source collection scheme including wearable physiological signals, video human postures and environmental sensors was designed, and an early feature fusion strategy based on local confidence adaptation was proposed. By dynamically adjusting the weights of each data source, it effectively dealt with the time-varying characteristics of data source contributions in different sports stages. At the feature extraction level, in response to the limitation of traditional time series models in simultaneously capturing multi-scale risk patterns, a multi-scale time series convolution network MHA-MSTCN was proposed, which parallelly fused four dilation rates of convolution features, and combined with a residual contraction module to achieve adaptive noise suppression related to the sports task, significantly improving the signal-to-noise ratio and discrimination ability of risk features. At the risk discrimination architecture level, a dual-path heterogeneous fusion model parallelly combining graph attention network and gated recurrent unit was designed, modeling the joint coupling abnormalities of human motion chain and the temporal drift of physiological parameters, and further achieving collaborative perception of spatial abnormalities and physiological abnormalities through cross-path cross-attention mechanism, enabling the model to capture the combined risk precursors of joint compensation and heart rate variation. At the warning mechanism level, a dynamic risk level division method based on K-means++ unsupervised clustering and contour coefficient was proposed, combined with individualized historical baseline to achieve threshold adaptive adjustment, and a multi-level rule engine with delay tolerance and hysteresis characteristics was designed, effectively balancing the sensitivity and stability of the warning. Experimental results show that the model achieved 92.3% accuracy, 89.7% recall rate and 97.1% AUC value on the test set, with an early warning time of 2.31 seconds, significantly superior to baseline models such as LSTM, CNN and Transformer. Abolition experiments verified the effective contributions of multi-head attention, multi-scale structure, dual-path architecture and cross-attention modules. Robustness tests showed that the model still maintained 92.7% AUC under 20% data packet loss conditions, and the lightweight design made the single-frame inference delay as low as 7.6 milliseconds and the model parameter size as only 1.42×10^5, meeting the requirements for real-time deployment at the edge end. The analysis of three typical cases further visually demonstrated the warning timing and accuracy of the model in actual sports scenarios.

7.2 Model limitations and future improvement directions

Although the model in this paper achieved relatively ideal performance under experimental conditions, there are still several limitations that need to be improved in future work. Firstly, the current model relies on a relatively complete multi-source sensor configuration, including wrist wearable devices, high-speed cameras, and embedded ground pressure sensors, which makes the deployment cost high and limits the model's large-scale promotion in ordinary primary and secondary school physical education classes. Although this paper has carried out lightweight design, there is still a dependence on hardware. Future work will explore knowledge distillation from multi-source input to sparse input and investigate whether posture information can be reconstructed using temporal generation models while retaining only wearable devices (heart rate + accelerometer) and basic subject information (age, exercise habits, physical fitness levels). Secondly, the training data of the model comes from 120 students of a single middle school. Although the sample size is relatively large, the diversity of the subjects' ages, exercise habits, and physical fitness levels is still limited. The movement patterns and physiological baselines of students from different regions and different age groups may have significant differences, and the generalization ability of the model in cross-scenario migration has not been fully verified. Future work will conduct multi-center, large-scale data collection, covering primary, junior high, and senior high school students in different grades and various climate conditions and sports field types, and study the rapid transfer method of the model based on domain adaptation, such as using adversarial domain alignment or unsupervised domain adaptation techniques, so that the model can maintain good early warning performance even when there is only a small amount of labeled or even unlabeled data in the target scenario. Thirdly, the current model uses offline batch training methods. Once deployed, the model parameters are fixed, and it is difficult to continuously optimize for individual students' adaptive changes during long-term exercise. Each student's physical fitness level, exercise skills, and injury susceptibility will drift with training accumulation, and the static model cannot capture this personalized evolution trend. Future work will explore online incremental learning frameworks and design lightweight model update mechanisms, so that the model can continuously update the model with small steps using real-time data streams generated at the edge, while avoiding catastrophic forgetting problems. Fourthly, although the warning rule engine currently has multi-level and delay tolerance characteristics, its parameters (such as the sliding window length, the number of continuous confirmations) are fixed values set through experimental experience and have not yet achieved closed-loop adaptive adjustment with the dynamic evolution of individual risks. Future work can introduce reinforcement learning frameworks, model the warning trigger strategy as a sequential decision-making problem, and use the cost of false alarms and missed alarms as reward signals, so that the warning rules can be automatically optimized based on real-time feedback. Fifthly, this paper mainly focuses on acute risk warnings during the exercise process and has not covered delayed injury risks during the recovery period (such as stress fractures caused by overtraining). Future work can expand the research perspective from real-time warnings during exercise to recovery assessment and long-term load management, and build a complete closed-loop system from single exercise risk warnings to periodic training load optimization. Finally, in terms of human-machine collaboration, the current model only outputs warning signals and has not been deeply integrated with interactive interfaces for teachers or students. Future work will develop a complementary mobile terminal visualization system, not only displaying risk levels, but also highlighting abnormal joint positions through heat maps, showing the evolution trajectory of physiological indicators through trend charts, and providing interpretable risk attribution analysis, so that the recipient can understand why the model issues a warning, thereby enhancing the trust and response efficiency of the warning signal. In conclusion, this paper provides a feasible technical framework for student sports risk warning based on multi-source data fusion. Future research will focus on reducing the deployment threshold, enhancing cross-scenario generalization ability, achieving personalized continuous evolution, and ultimately promoting this technology from the laboratory to real sports education scenarios.

Abbreviations

TCN, Temporal Convolutional Network;

MHA-MSTCN, Multi-Head Attention Multi-Scale Temporal Convolutional Network;

GAT, Graph Attention Network;

GRU, Gated Recurrent Unit;

CNN, Convolutional Neural Network;

LSTM, Long Short-Term Memory;

Transformer, Transformer;

ROC, Receiver Operating Characteristic;

AUC, Area Under the Curve;

FAR, False Alarm Rate;

TP, True Positive;

TN, True Negative;

FP, False Positive;

FN, False Negative;

WAT, Warning Advance Time;

NTP, Network Time Protocol;

OpenPose, Open Pose Estimation;

ReLU, Rectified Linear Unit;

ELU, Exponential Linear Unit;

GAP, Global Average Pooling;

INT8, 8-bit Integer;

KNN, K-Nearest Neighbors;

RMSprop, Root Mean Square Propagation;

AdamW, Adaptive Moment Estimation with Weight Decay;

GPU, Graphics Processing Unit;

CPU, Central Processing Unit;

RAM, Random Access Memory;

FPS, Frames Per Second.

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's contributions are specified as follows: **M.Z.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **H.L.:** Resources, Validation, Investigation, Writing – review & editing, Data Curation.

Funding information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data that support the findings of this study are available upon request from the corresponding authors, **H.L.**

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used DeepSeek for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

[1] Al-Qahtani, M. A., Allajhar, M. A., Alzahrani, A. A., Asiri, M. A., Alsalem, A. F., Alshahrani, S. A., ... & Alsalem, A. F. F. (2023). Sports-related injuries in adolescent athletes: a systematic review. *Cureus*, *15*(11). https://doi.org/10.7759/cureus.49392

[2] Zeng, Y., Ji, W., Shi, Y., Liu, W., & Ji, W. (2025). Sports injuries in elite football players: Classification, prevention, and treatment strategies update. *Frontiers in Sports and Active Living*, *7*, 1643789. https://doi.org/10.3389/fspor.2025.1643789

[3] Nair, M. J., & Swarna Nantha, Y. (2025). Common Sports, Muscular, and Ligamentous Injuries. In *Screen, Diagnose, Manage and Monitor: A GP Primer to Common Clinical Conditions* (pp. 1275-1283). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-7640-5_146

[4] Sixt, J. C. (2026). *High School Baseball Coaches Understanding and Management of Potential Mild Traumatic Brain Injuries in Catchers* (Doctoral dissertation, Grand Canyon University).

[5] Porsanger, L. (2023). Risk and safety management in physical education: teachers' knowledge. *Physical Education and Sport Pedagogy*, *28*(1), 16-28. https://doi.org/10.1080/17408989.2021.1934663

[6] Zou, X., Wang, L., & Han, R. (2025). Embedded Cyber-Physical Systems for Monitoring and Optimization of Sports Facilities Using Consumer IoT (CIoT) and Edge Computing. *IEEE Transactions on Consumer Electronics*, *71*(2), 4838-4848. https://doi.org/10.1109/tce.2025.3535162

[7] Jia, Y. (2026). Integrating Heart Rate Variability and Exercise Load for Early Warning of Overtraining Syndrome. *International Journal of Computer Information Systems and Industrial Management Applications*, *18*, 12-12. https://doi.org/10.70917/ijcisim-2026-1827

[8] Urbina, R. (2022). Utilizing Heart Rate as an Exercise Intensity Training Method to Determine Overtraining for Collegiate Long-Distance Runners. *SACAD: Scholarly Activities*, *2022*(2022), 12. https://doi.org/10.58809/gqki2597

[9] Shariatzadeh, M., Hafshejani, E. H., Mitchell, C. J., Chiao, M., & Grecov, D. (2023). Predicting muscle fatigue during dynamic contractions using wavelet analysis of surface electromyography signal. *Biocybernetics and Biomedical Engineering*, *43*(2), 428-441. https://doi.org/10.1016/j.bbe.2023.04.002

[10] Sarvestan, J., & Fakhraei Rad, N. (2025). Lower Limb Joint Coordination and Coordination Variability During Landing: A Scoping Review. *Applied Sciences*, *15*(9), 5118. https://doi.org/10.3390/app15095118

[11] Khosravi, M. R. (2025). Sports Injuries: Biomechanical Data Analysis and Prevention. *Advanced Journal of Management, Humanity and Social Science*, *1*(9), 559-578. https://doi.org/10.23977/JHMS.2025.060106

[12] Moreira, R., Fialho, R., Teles, A. S., Bordalo, V., Vasconcelos, S. S., de Morais Gouveia, G. P., ... & Teixeira, S. (2022). A computer vision-based mobile tool for assessing human posture: A validation study. *Computer methods and programs in biomedicine*, *214*, 106565. https://doi.org/10.1016/j.cmpb.2021.106565

[13] Wang, X. (2024). Biomechanical Analysis on the In-situ Three-point Jump Shooting Technique in Basketball. In *Sustainable Materials Processing and Manufacturing* (pp. 1-6). CRC Press. https://doi.org/10.1201/9781003538646-1

[14] Yan, W., Jiang, X., & Liu, P. (2023). A review of basketball shooting analysis based on artificial intelligence. *IEEE Access*, *11*, 87344-87365. https://doi.org/10.1109/access.2023.3304631

[15] Mason, R., Pearson, L. T., Barry, G., Young, F., Lennon, O., Godfrey, A., & Stuart, S. (2023). Wearables for running gait analysis: A systematic review. *Sports Medicine*, *53*(1), 241-268. https://doi.org/10.1007/s40279-022-01760-6

[16] Man, K., & Chahl, J. (2022). A review of synthetic image data and its use in computer vision. *Journal of Imaging*, *8*(11), 310. https://doi.org/10.3390/jimaging8110310

[17] Jaiswal, D., & Kumar, P. (2022). A survey on parallel computing for traditional computer vision. *Concurrency and Computation: Practice and Experience*, *34*(4), e6638. https://doi.org/10.1002/cpe.6638

[18] Oyeleye, M., Chen, T., Titarenko, S., & Antoniou, G. (2022). A predictive analysis of heart rates using machine learning techniques. *International Journal of Environmental Research and Public Health*, *19*(4), 2417. https://doi.org/10.3390/ijerph19042417

[19] Ghafoori, M., Clevenger, C., Abdallah, M., & Rens, K. (2023). Heart rate modeling and prediction of construction workers based on physical activity using deep learning. *Automation in Construction*, *155*, 105077. https://doi.org/10.1016/j.autcon.2023.105077

[20] Gu, M., Zhang, Y., Wen, Y., Ai, G., Zhang, H., Wang, P., & Wang, G. (2023). A lightweight convolutional neural network hardware implementation for wearable heart rate anomaly detection. *Computers in Biology and Medicine*, *155*, 106623. https://doi.org/10.1016/j.compbiomed.2023.106623

[21] Canessa-Pollard, V., Anikin, A., & Reby, D. (2025). Chants across seven traditions share acoustic traits that enhance subjective relaxation. *Proceedings of the National Academy of Sciences*, *122*(35), e2506480122. https://doi.org/10.1073/pnas.2506480122

[22] Yunardi, R. T., Sardjono, T. A., & Mardiyanto, R. (2025). Enhancing Surveillance Vision-Based Human Action Recognition Using Skeleton Joint Swing and Angle Feature and Modified AlexNet-LSTM. *International Journal of Intelligent Engineering & Systems*, *18*(1). https://doi.org/10.22266/ijies2025.0229.53

[23] Chen, H., & Fan, R. (2025). Improved convolutional neural network for precise exercise posture recognition and intelligent health indicator prediction. *Scientific Reports*, *15*(1), 21309. https://doi.org/10.1038/s41598-025-01854-x

[24] Alif, F. K. (2026). Extending peaks-over-threshold methods to public health: modified threshold selection and scenario-based risk estimation. *Health Services and Outcomes Research Methodology*, 1-25. https://doi.org/10.21203/rs.3.rs-7553933/v1

[25] Liu, P. L., & Chang, C. C. (2022). Simple method integrating OpenPose and RGB-D camera for identifying 3D body landmark locations in various postures. *International Journal of Industrial Ergonomics*, *91*, 103354. https://doi.org/10.1016/j.ergon.2022.103354

[26] Balakrishnan, K., Dhanalakshmi, R., & Gopalakrishnan, R. (2023). Clock synchronization in industrial Internet of Things and potential works in precision time protocol: Review, challenges and future directions. *International Journal of Cognitive Computing in Engineering*, *4*, 205-219. https://doi.org/10.1016/j.ijcce.2023.06.001

[27] Xie, Q., Tao, G., Xie, C., & Wen, Z. (2022). Abnormal data detection based on adaptive sliding window and weighted multiscale local outlier factor for machinery health monitoring. *IEEE Transactions on Industrial Electronics*, *70*(11), 11725-11734. https://doi.org/10.1109/tie.2022.3231279

[28] Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. *ACM computing surveys*, *56*(9), 1-36. https://doi.org/10.1145/3649447

[29] Pant, A., & Kumar, A. (2024). Hanning FIR window filtering analysis for EEG signals. *Biomedical Analysis*, *1*(2), 111-123. https://doi.org/10.1016/j.bioana.2024.05.003

[30] Ayodeji, A., Wang, Z., Wang, W., Qin, W., Yang, C., Xu, S., & Liu, X. (2022). Causal augmented ConvNet: A temporal memory dilated convolution model for long-sequence time series prediction. *ISA transactions*, *123*, 200-217. https://doi.org/10.1016/j.isatra.2021.05.026

[31] Giuste, F. O., He, L., Lais, P., Shi, W., Zhu, Y., Hornback, A., ... & Wang, M. D. (2023). Early and fair COVID-19 outcome risk assessment using robust feature selection. *Scientific Reports*, *13*(1), 18981. https://doi.org/10.1038/s41598-023-36175-4

[32] Vrahatis, A. G., Lazaros, K., & Kotsiantis, S. (2024). Graph attention networks: a comprehensive review of methods and applications. *Future Internet*, *16*(9), 318. https://doi.org/10.3390/fi16090318

[33] Zhang, Q., & Qian, J. (2025). Identification and Temporal distribution of typical rainfall types based on K-Means++ Clustering and probability distribution analysis. *Hydrology*, *12*(4), 88. https://doi.org/10.3390/hydrology12040088

APA Citation

Zhou, M., & Lei, H. (2026). Construction of a Student Sports Risk Early Warning Model Based on Multi-Source Data Fusion. *Journal of Digital Frontier, 1*(1), 31-60. **https://doi.org/10.67541/JDF2602**