CNN and LSTM-Based Multimodal Data Fusion for Performance Optimization in Aerobics Using Wearable Sensors

Danhua Tan
School of Physical Education, Hengyang Normal University, Hengyang, Hunan, 421006, China
E-mail: tandanhua184914@outlook.com

Keywords: wearable sensors, convolutional neural network, long short-term memory, kalman filtering, aerobics movements

Received: May 31, 2025

Aerobics is a high-intensity, multi-dimensional sport. Its motion evaluation places higher demands on data quality and time series modeling capabilities. This paper proposes a method for evaluating aerobics motion that integrates wearable sensors and motion tracking systems. It combines convolutional neural networks (CNNs) with long short-term memory networks (LSTMs) to perform fusion analysis on multimodal data from accelerometers, gyroscopes, magnetometers, and Kinect motion capture systems. To improve data quality, Kalman filtering, time synchronization, and wavelet transform techniques are introduced to preprocess the raw data. Experimental results show that this method performs well in motion classification tasks: in indoor low-intensity training scenarios, the accuracy of the CNN model increases from 74.5% to 87.1%; in high-intensity training scenarios, the accuracy increases from 75.0% to 88.2%. After combining with LSTM, the model further enhances the modeling capabilities of motion time series features and improves the recognition accuracy of complex motions. In different training scenarios, the average improvement rate of motion scores is 25.8%. The system feedback delay is controlled within 200 milliseconds, with good real-time and practical performance. This method provides aerobics athletes with high-precision movement assessment and personalized training suggestions, promoting the intelligent and personalized development of sports training.

Povzetek: Metoda združuje senzorje, CNN in LSTM za multimodalno analizo aerobičnih gibov. Kalmanovo filtriranje izboljša kakovost signalov, klasifikacijska točnost naraste do 88,2 %, povprečno izboljšanje rezultatov znaša 25,8 %, odzivnost sistema pa ostane pod 200 ms.

1 Introduction

With the popularity of aerobics, the accuracy of movements and training effects have become the focus of coaches and athletes. During high-intensity and complex exercise, the movements of aerobics athletes can be affected by factors such as physical exertion, sports skills, and external environment, resulting in unstable movement performance. Traditional manual evaluation methods are inefficient and subjective, and cannot provide athletes with accurate training feedback in real-time. With the development of sensor technology [1] and artificial intelligence [2], [3], motion evaluation methods based on wearable devices [4] and intelligent feedback systems have become a research hotspot. Such feedback systems can provide accurate real-time data analysis, optimize training programs, and improve athlete performance. Therefore, developing a motion evaluation and optimization system based on intelligent technology [5], [6] has become the key to improving training effects and athlete performance.

This paper studies the motion evaluation and optimization system based on intelligent technology to improve the training effect and motion performance of aerobics athletes. To achieve this goal, this paper combines wearable sensors with motion tracking systems and uses CNN models and LSTM to fuse and analyze multimodal data. The system acquires motion data through sensors such as accelerometers, gyroscopes, magnetometers, and Kinect motion capture systems. It uses Kalman filtering, time synchronization, and wavelet transform to optimize data quality. The optimized data is used through the CNN model to evaluate and optimize motion performance, providing real-time feedback and personalized training suggestions. The CNN-based optimization method combines wearable sensor technology with deep learning (DL) algorithms to improve the accuracy and stability of motion evaluation. Experimental results show that the combination of Kalman filtering and CNN models effectively improves the accuracy and stability of aerobics motion evaluation, providing strong support for the intelligent and precise development of sports training.

Current research mostly uses weighted averaging or simple concatenation, and the model structure is fixed, without optimizing for the temporal characteristics and complex action patterns of sports data. This paper combines wearable sensors with aerobics tracking and uses a model based on CNN and LSTM to achieve performance optimization. The main contributions of this

study include: 1) wavelet transform is combined and principal component analysis (PCA) to extract time-frequency features, and a dynamic weighted fusion strategy is adopted to improve the robustness of data fusion; 2) small convolutional kernels are introduced into CNN to capture action details and combined with double-layer LSTM to model long-term dependencies, enhancing the model's ability to recognize complex action sequences; 3) based on the model output, an action scoring function and error correction mechanism are constructed to provide athletes with immediate feedback and personalized training suggestions, improving the model's generalization ability in different training scenarios through data augmentation and adaptive filtering techniques.

2 Related work

In recent years, many scholars have been committed to improving the accuracy of athletes' motion evaluation through different technical means. Traditional motion capture systems [7] rely too much on calibration equipment and high-cost hardware settings. Although such capture systems can capture the movements of athletes, they suffer from problems such as poor real-time performance, high data noise, and inconvenient operation when evaluating high-intensity sports or complex movements. To improve the quality of sports data, many studies have attempted to use wearable sensors for motion tracking. Rigozzi C J et al. used data from sensors such as accelerometers, gyroscopes, and magnetometers to monitor athletes' body posture and motion trajectory [8]. Sensor data is easily affected by noise, environmental changes, and wear position deviation, resulting in inaccurate data. To reduce noise interference, Zhang Y applied Kalman filtering technology to the preprocessing of sensor data [9]. As technology matures, DL technology [10], especially CNNs, has been applied to multimodal data analysis and action recognition by Gholamiangonabadi D [11], and has achieved certain results. Existing research still faces problems such as how to combine multiple data sources, optimize data processing processes, and provide real-time feedback in actual training scenarios.

To solve the above problems, some researchers have proposed a hybrid method that combines sensor data and DL algorithms to improve the accuracy and real-time performance of action recognition. Chakraborty A used a CNN-based multimodal data fusion method to improve the accuracy and robustness of athlete action recognition by combining accelerometer, gyroscope, and visual data [12]. In this study, data fusion technology [13] effectively reduces sensor errors and enhances the system's adaptability to complex actions. Zhang L proposed a KCF (Kernelized Correlation Filters) tracking method based on improved depth information [14], which successfully used Kalman filtering to reduce the noise of motion sensors and improve the stability of motion estimation. Although these methods have achieved good results to a certain extent, most of them focus on a single motion estimation task, and their effects in complex training environments still need to be improved. Existing methods also have shortcomings in terms of personalized training feedback [15] and the generation of real-time optimization suggestions [16]. Therefore, how to comprehensively utilize multimodal data and combine DL with real-time optimization feedback systems is still a major challenge in current research.

3 Data fusion and movement performance optimization

3.1 Data collection and preprocessing

The study combines wearable sensors with motion tracking systems to design an efficient data collection and preprocessing solution. The key to the entire process is to synchronously collect data from multiple sources and eliminate errors, providing a reliable basis for subsequent analysis.

Wearable devices collect data in real-time through built-in accelerometers [17], gyroscopes [18], and magnetometers [19]. The accelerometer records the athlete's acceleration changes in three-dimensional space; the gyroscope measures the athlete's rotational angular velocity; the magnetometer helps correct the direction of movement. A multi-sensor system can accurately capture every movement of an athlete and generate rich time series data [20]. Sensor data fusion equation is:

$$f(t) = \alpha \cdot a(t) + \beta \cdot \omega(t) + \gamma \cdot m(t) \quad (1)$$

$a(t)$ is acceleration data; $\omega(t)$ is angular velocity data; $m(t)$ is magnetic field data; α, β, γ are weighting parameters. Figure 1 is a data acquisition flow chart.

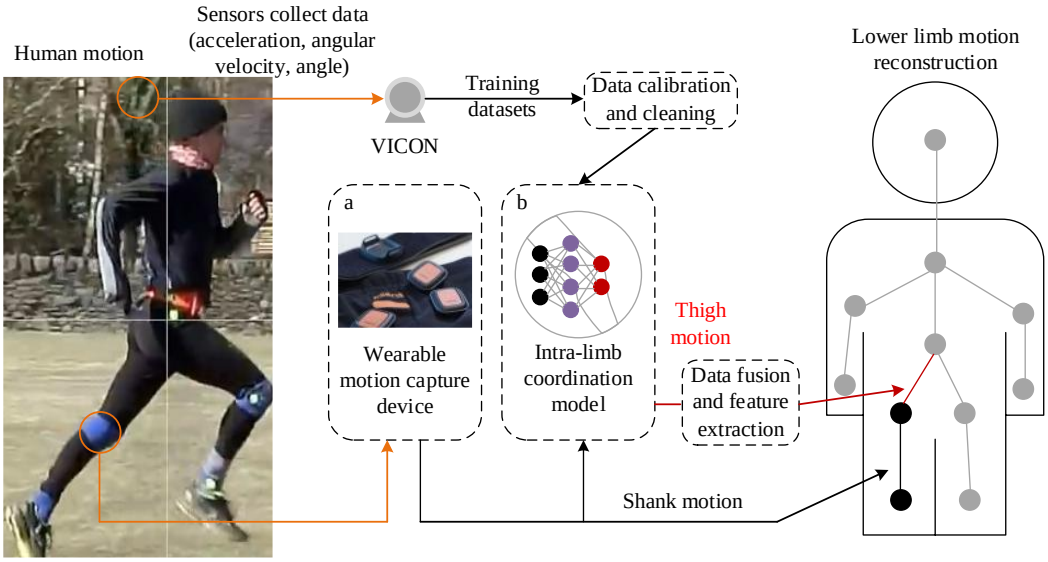


Figure 1: Data collection flow chart

Figure 1 shows the complete process from data collection to motion analysis. The motion data is obtained through wearable sensors and motion capture systems; key features are extracted through data cleaning and fusion; the CNN model is used to optimize the motion performance evaluation, ultimately providing scientific motion optimization and training suggestions for aerobics athletes. To ensure the accuracy of the data, all collected sensor data are transmitted to the background data processing system in real-time via Bluetooth or Wi-Fi modules [21] to ensure the real-time and efficient data. By adopting wireless data transmission [22], the data transmission is not affected by the physical distance, ensuring that the data is updated and recorded in time when the athletes perform complex movements. Bluetooth signal quality function is:

$$Q = \frac{1}{1 + \exp(-k(S - S_0))} \tag{2}$$

S is the signal strength; S_0 is the signal threshold; k is the Bluetooth signal adjustment parameter. To deal with data anomalies, the Kalman filter [23] is used to smooth the data of accelerometers and gyroscopes. The Kalman filter can dynamically predict the true value of the signal, optimize the measurement noise, and improve the accuracy of the data. Kalman filter formula is:

$$x_{k|k} = x_{k|k-1} + K_k(z_k - H_{x_{k|k-1}}) \tag{3}$$

K_k is the Kalman gain, and z_k is the observed value.

In addition to multi-sensor equipment, the motion tracking system Kinect [24] and depth camera [25] are also introduced to obtain the spatial position information of the key parts of the athletes. The system captures the athlete's action posture through 3D coordinates and records the spatial coordinates of joints such as shoulders, elbows, and knees, as well as their dynamic trajectories over time. Through the calibration algorithm, combined with the position information of the sensor and the motion tracking system, the effects caused by the wearer position offset or motion capture error are corrected. Motion trajectory smoothing formula is:

$$p(t) = \frac{1}{N}\sum_{i=1}^{N} p_i(t) \tag{4}$$

$p_i(t)$ is the spatial position of different sampling points. The system synchronizes the data of sensors and motion tracking systems to ensure that the sensor data and motion data at each moment can correspond correctly. After time synchronization, the data can be smoothly input into the subsequent data processing and analysis. Time synchronization function is:

$$\Delta T = T_{sensor} - T_{camera} \tag{5}$$

T_{sensor} and T_{camera} are the timestamps of the sensor and camera, respectively. Table 1 is the motion capture key point coordinate data table.

Table 1: Motion capture key point coordinate data table.

Timestamp (ms)	Shoulder X (cm)	Shoulder Y (cm)	Shoulder Z (cm)	Knee X (cm)	Knee Y (cm)	Knee Z (cm)
0	12.3	45.6	78.2	8.9	30.2	50.7
10	12.1	45.5	78.3	9	30.3	50.9
20	12.2	45.7	78.1	9.1	30.1	50.8
30	12.4	45.8	78.2	9.2	30.4	51
40	12.3	45.9	78.3	9.3	30.5	50.9
50	12.5	46	78.1	9.4	30.6	51.1
60	12.6	46.1	78.2	9.5	30.7	51.2

Table 1 records the three-dimensional spatial position data (X, Y, Z, in centimeters) of the shoulder and knee at different timestamps (in milliseconds). The data can be used to analyze the movement trajectory and change trend of the shoulder and knee in space.

3.2 Multimodal data fusion

After data collection and preprocessing, the multimodal data from wearable sensors and motion tracking systems are effectively fused. Different data sources provide different perspectives on the athlete's movements. Wearable sensor data provides time series information such as acceleration and angular velocity, and the motion tracking system provides spatial information such as joint position and motion trajectory. The effective integration of this information can help comprehensively evaluate the athlete's performance and provide accurate data input for the DL model.

Time synchronization [26] is a prerequisite for ensuring the effective integration of multimodal data. When collecting sensor data and motion tracking data, the data needs to be accurately aligned in time due to the different collection frequencies of the two [27]. To achieve data calibration, a timestamp is used to mark the acquisition time of each frame of data to ensure that each action frame can obtain corresponding sensor data and tracking data. After time synchronization, the sensor data at each moment is guaranteed to correspond perfectly with the action tracking data, providing a basis for data fusion. After synchronization correction, it can ensure that the action and sensor data at each moment correspond to each other, avoiding information loss caused by asynchrony [28].

The original sensor data contains rich time series information. Wavelet transform [29] is used to analyze the data in the time and frequency domain to extract motion features. Wavelet transform formula is:

$$W_{\psi}(a,b) = \int_{-\infty}^{\infty} f(t)\psi^{*}(\frac{t-b}{a})\,dt \tag{6}$$

ψ is the mother wavelet function. Wavelet transform can effectively capture the instantaneous changes in motion signals and use multi-scale analysis [30] to extract the time-frequency features of motion signals. It is integrated with the sensor data by calculating the spatial characteristics of the athlete's joint angle, motion trajectory, and speed. The formula for calculating the joint angle is:

$$\theta = arc\,cos(\frac{v_1 \cdot v_2}{\|v_1\|\|v_2\|}) \tag{7}$$

v_1 and v_2 are the vectors of two bones. The motion trajectory curve fitting formula is:

$$r(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n \tag{8}$$

The spatial characteristics of the tracking system play a decisive role in the accuracy and coordination of the movements and are the core basis for evaluating the performance of athletes. Data fusion is the key to multimodal data processing. When fusion is performed, methods such as weighted fusion [31] and principal component analysis (PCA) [32] are used. Weighted fusion is to assign different weights to different data sources according to their signal-to-noise ratio and importance. When the noise of sensor data is large, its weight in fusion is reduced. On the contrary, if it is small, it means that the spatial data provided by the motion tracking system is relatively stable and can be assigned a higher weight. Weighted fusion formula is:

$$F_{fused} = \omega_1 F_1 + \omega_2 F_2 \tag{9}$$

F_1 and F_2 are the features of different data sources, and ω_1 and ω_2 are weights. Through weighted fusion processing, the fused data can more realistically reflect the athlete's performance. In weighted fusion, the quality of the signal is the key to weight assignment, and the relevance and accuracy of the data determine the contribution of each source information.

Principal component analysis is used to reduce the dimensionality of the data, compressing the multi-dimensional raw data into fewer principal components, reducing the redundancy of the data, and extracting the most representative features. PCA dimensionality reduction formula is:

$$X' = XW, W = arg\ \max_{W}\ (\frac{W^T \sum W}{W^T D W}) \tag{10}$$

Σ is the feature covariance matrix [33]; D is the weight matrix [34]; W is the optimized feature matrix. The analysis process reduces the computational complexity and retains the key information in the data, providing more efficient input for the training of DL models. Through the dimensionality reduction of PCA, redundant dimensions and noise can be eliminated, improving the efficiency and accuracy of subsequent analysis.

After fusion processing of multimodal data, data from different sources is integrated into a unified format, providing rich and accurate input features for subsequent action evaluation. The fused data contains both time series information and spatial position information, which can fully and accurately reflect the athlete's action performance in training [35]. Combined with efficient data synchronization, feature extraction, and fusion processing, sufficient high-quality data support is provided for the CNN, ensuring that the model can make full use of various types of information for accurate evaluation.

3.3 Action performance evaluation based on CNN and LSTM

This paper studies the evaluation of action performance of fused data based on CNN and LSTM. The CNN model has an advantage in processing time series data and spatial data, while LSTM is good at capturing time series dependencies, especially in capturing subtle differences in athletes' movements and automatically extracting features. The CNN model can more comprehensively evaluate the athletes' movement performance and achieve end-to-end automated processing from raw sensor and tracking data to final movement scoring and classification. Through LSTM, the model can understand the continuity between actions and evaluate actions based on the relationship

between the action sequence. When capturing complex motion patterns, LSTM can supplement the timing information that the CNN model fails to fully capture, providing a more detailed motion performance evaluation. In the motion performance evaluation task, the combination of LSTM and CNN enables the model to obtain more comprehensive feature extraction in both spatial and temporal dimensions. After CNN extracts spatiotemporal features, LSTM processes these features in time series, and the combination can more accurately evaluate the quality and type of actions. Figure 2 is a diagram of the CNN model structure.

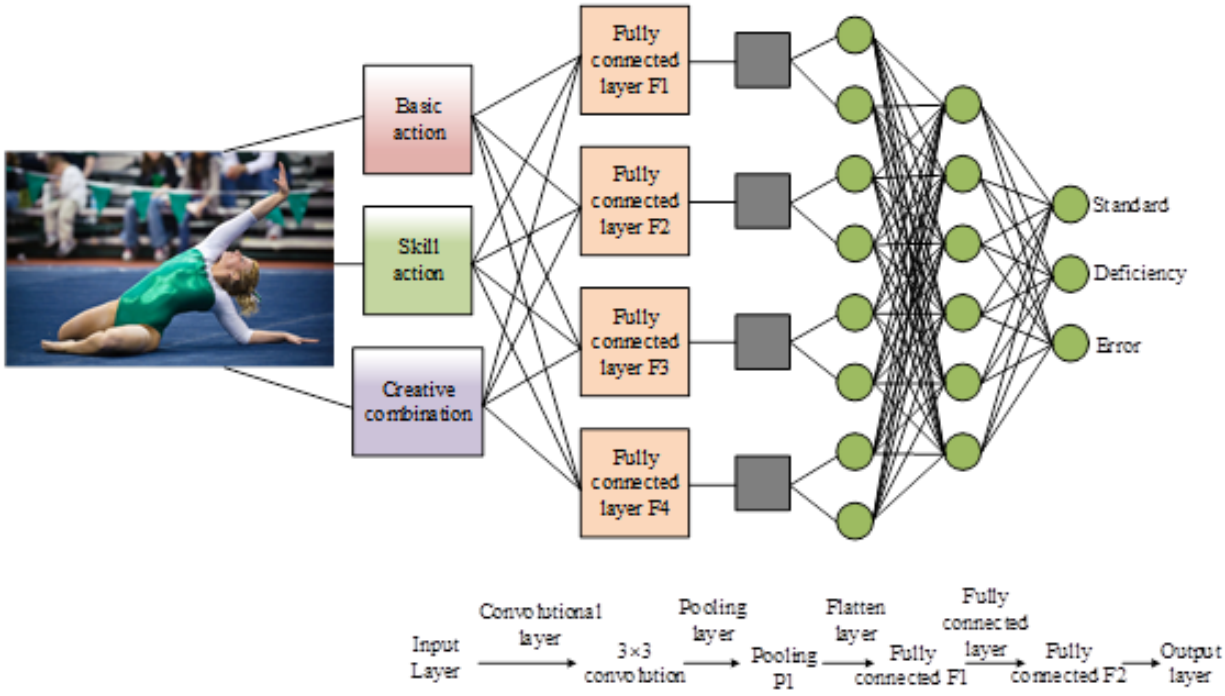


Figure 2: CNN model structure diagram

Figure 2 shows how the spatiotemporal features of the athlete's movements are gradually extracted through the convolutional layer, pooling layer, and fully connected layer, and converted into a one-dimensional vector for classification and scoring through the flattening layer. The output layer evaluates the athlete's movement quality based on the features learned by the model, achieving automated and efficient movement quality recognition and feedback.

The CNN model adopts a five-layer structure, with input data dimensions of (T, F). Among them, T=100 represents the time step; F=9 represents the input feature dimension (including data from three axes each of accelerometer, gyroscope, and magnetometer); the output dimension is (C), where C=3 represents the action category (standard, insufficient, error); the activation function is Softmax. The LSTM model is used to model the long-term dependencies of time series, with input dimensions of (T, D), where T=100 represents the time step, and D=9 represents the feature dimension. The LSTM layer contains 128 hidden units and uses a double-layer stacking structure with an activation function of Tanh. The output dimension is (C), C=3, and the activation function is Softmax. The output feature vectors of CNN and LSTM are merged through concatenation operation and input into a fused fully connected layer, ultimately outputting action scores and classification results. The model parameter settings are shown in Table 2:

Table 2: Model parameter settings

Parameter	Specifications	Parameter	Specifications
Learning rate	0.001	Dropout rate	0.5
Batch size	64	Training epochs	100
Optimizer	Adam	Hidden layer size	128
Loss function	Cross entropy loss function	Convolutional kernel size	(3, 3)

The convolution operation can effectively capture the spatial features and temporal dynamic changes of the athlete's joint movement trajectory, body posture changes, etc. The features processed by CNN can be passed to the LSTM network, which further analyzes the timing information of these features. Through the time memory mechanism of LSTM, the network can learn the continuity and long-term dependencies in the action. Convolution operation formula is:

$$f_{i,j} = \sum_{m=-k}^{k} \sum_{n=-k}^{k} x_{i+m,j+n} \cdot w_{m,n} \tag{11}$$

x is the input feature map; w is the convolution kernel; the choice of the convolution kernel is closely related to the characteristics of the data. Smaller convolution kernels help capture subtle changes, while larger convolution kernels can extract more macro features. It is necessary to adjust the bodybuilder's tiny movements and postures, and smaller convolution kernels help extract details more accurately, so a convolution kernel of size 3×3 is selected. The LSTM layer filters out irrelevant temporal information through its forget gate, input gate, and output gate, retaining the long-term and short-term dependency information related to the action performance evaluation.

After the convolution layer, the maximum pooling and average pooling [36] are used to reduce the dimension of the feature map. The role of the pooling layer is to reduce the amount of data after the convolution operation, reduce the computational complexity, and retain the most important feature information. Pooling operation formula is:

$$p_{i,j} = \underset{m,n}{max} (g_{i+m,j+n}) \tag{12}$$

In the fully connected layer, the temporal information generated by LSTM and the spatial features extracted by CNN are integrated through the fully connected layer to generate the final score and classification results of the action performance. The model can not only evaluate the quality of actions, but also classify actions into multiple categories such as "standard", "deficient", and "error". Action scoring function is:

$$S = \sum_{i=1}^{N} \varphi_i h_i \tag{13}$$

h_i is the score of each feature, and φ_i is the weight. The action evaluation score reflects the accuracy, fluency, and standardization of the athlete's action. The classification results provide coaches and athletes with targeted training improvement directions. Each classification result helps to further guide athletes' specific improvement measures in training.

During the training process, the Adam optimization algorithm [37] is used to update parameters. The optimization algorithm update formula is:

$$\theta_{t+1} = \theta_t - \eta \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \tag{14}$$

$\widehat{m}_t$ and $\widehat{v}_t$ are momentum estimates. The cross-entropy loss function [38] is used to optimize the classification task, so that the model can better handle multi-classification problems and continuously improve the accuracy of action scoring and classification by minimizing the loss function. Classification loss function is:

$$L = -\sum_{i=1}^{C} y_i \, log(\hat{y}_i) \tag{15}$$

y_i is the true label, and $\hat{y}_i$ is the predicted probability. To improve the model's training effect, data enhancement technology is used to simulate the motion performance in different training scenarios, expand the training data set, and increase the robustness of the model. Data enhancement methods include operations such as rotation, mirror flipping, and scaling. More training samples are generated through enhancement operations, so that the model can still show excellent performance in different motion modes.

3.4 Real-time feedback and optimization suggestion generation

To improve the effect of aerobics training, a real-time feedback system is designed to feed back the evaluation results generated during the training process to athletes, helping them adjust their movements, avoid errors and optimize the quality of movements. The key to the feedback system lies in real-time and accuracy. Only timely and accurate feedback can effectively improve the training level of athletes.

The system analyzes the real-time action data of athletes through a model combining a trained CNN model and LSTM to generate more accurate action scores and evaluation results. Every time an athlete performs an action, the system immediately analyzes the action and outputs a real-time score. The score reflects the accuracy, fluency, and completion of the action. The higher the score, the more standard the action. For low-scoring actions, the system automatically identifies the errors and provides specific optimization suggestions. Error correction weight formula is:

$$\varrho_{error} = \frac{1}{1 + exp(-\kappa \cdot \delta)} \tag{16}$$

δ is the margin of error. Based on the score and error recognition results, the system generates personalized optimization suggestions. Optimization suggestion generation function is:

$$G = arg \underset{i}{max} (S_i + \lambda \cdot E_i) \tag{17}$$

S_i is the score, and E_i is the severity of the error. Suggestions include improving posture, adjusting the range of motion, strengthening muscle control, etc., to help athletes correct deficiencies in their movements. Optimization suggestions can be provided in text form or visualized through a graphical interface to help athletes more intuitively understand the problems in their movements.

To ensure timely feedback, the system accelerates the calculation process by optimizing the algorithm, controlling the delay between action scoring and feedback generation to less than 200ms, ensuring that athletes can receive targeted adjustment suggestions in a short period of time. Real-time feedback delay formula is:

$$T_{delay} = T_{process} + T_{transmit} \tag{18}$$

Through the real-time feedback mechanism, athletes can continuously adjust their movements during training and gradually improve the training effect. Optimization suggestions are not limited to correcting mistakes, but can also help athletes improve the delicacy and accuracy of their movements. The athlete improvement index formula is:

$$I = \frac{\Delta S \cdot ln(1 + A_0)}{\Delta T \cdot (1 + e^{-\varepsilon(\Delta S - \Delta S_{avg})})} \quad (19)$$

ΔS is the score increment; ΔT is the time; A_0 is the athlete's baseline ability; ε is the adjustment parameter; ΔS_{avg} is the average score increment. Through long-term training and optimization feedback, athletes can improve their overall performance in a short period of time and achieve the best training effect. Long-term optimization trend equation is:

$$O(t) = O_0 + \int_0^t \frac{dS}{d\tau} d\tau \quad (20)$$

O_0 is the initial performance. Combining DL with real-time feedback technology, it provides athletes with an intelligent training platform that can effectively improve training efficiency and quality. Table 3 is some hyperparameter data of the experiment.

Table 3: Some hyperparameter data

Parameters	Function	Parameters	Function
α, β, γ	Sensor data fusion weighting parameters	k	Bluetooth signal conditioning parameters
K_k	Kalman Gain	ψ	Mother wavelet function
ω_1, ω_2	Weighted fusion weight	Σ	Feature covariance matrix
w	Convolution Kernel	φ_i	Action score weight
$\hat{m}_t$, $\hat{v}_t$	Momentum Estimation	$\hat{y}_i$	Prediction probability
δ	Margin of Error	ε	Adjustment parameters

Table 3 lists the hyperparameters used for model and signal processing and their functional descriptions. These hyperparameters play a key role in sensor data fusion, signal conditioning, feature extraction, and prediction, and can effectively improve the performance and accuracy of the model. Adjusting these parameters can optimize system behavior according to specific application requirements and achieve more accurate data processing and action recognition.

4 Experimental results

4.1 Experimental setup

This paper collects a total of 12000 action samples, covering three categories of actions: standard actions (4000), insufficient actions (4000), and incorrect actions (4000). Each sample contains multimodal data of 100 time steps, including 9 channels (3-axis x 3 sensors) from accelerometers, gyroscopes, magnetometers, and 18 joint point 3D coordinate information obtained by Kinect. The data collection frequency is 50 Hz, which means collecting 50 frames per second. The wearable sensor is the Xsens MTw Awinda series inertial measurement unit (IMU), with specific parameters shown in Table 4:

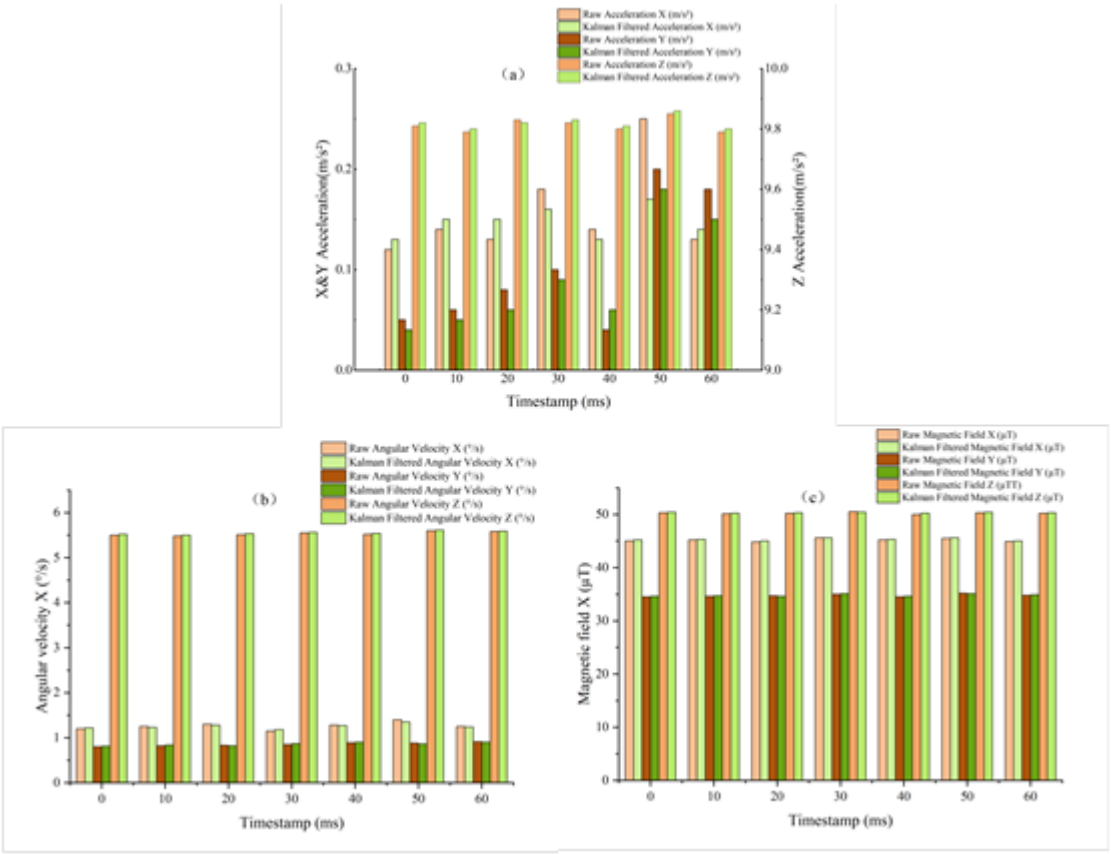
Table 4: Sensor parameters

Sensor	Range	Resolution	Sampling rate
Accelerometer	±16g	0.001g	50 Hz
Gyroscope	±2000°/s	0.01°/s	50 Hz
Magnetometer	±2.5 Gauss	0.01 Gauss	50 Hz

The data collection is conducted in an indoor sports arena, with an ambient temperature controlled at 22-25 ° C, humidity at 45-60%, and good and stable lighting conditions. All model training is completed under the PyTorch 1.13.1 deep learning framework, with a training time of approximately 4 hours per model.

4.2 Effect of Kalman Filtering on sensor data

In sensor data processing, the original signal is easily affected by noise, which affects the accuracy of the data. Kalman filtering, as a common noise suppression method, can effectively improve the stability and accuracy of the data by correcting the measured values. Figure 3 shows the comparison between the original data of the accelerometer, gyroscope, and magnetometer and the data after Kalman filtering, which is used to evaluate the filtering effect.



(a) Accelerometer data comparison; (b) Gyroscope data comparison; (c) Magnetometer data comparison

Figure 3: Kalman filter effect

Figure 3 shows the fluctuation of different sensors in different directions and the effect after filtering. In the accelerometer data, after Kalman filtering, the peak-to-valley difference of acceleration X drops from the original 0.13m/s² to 0.04m/s² after filtering, showing a more stable state. The original acceleration fluctuates greatly in the X-axis and Y-axis directions. After Kalman filtering, the fluctuation of the data is more stable, indicating that Kalman filtering can effectively reduce the interference of noise. When the timestamp is 0ms, the original acceleration X is 0.12m/s², and after Kalman filtering, it is 0.13m/s², with little change. The gyroscope data also shows a similar trend. The original data fluctuates to varying degrees in the X, Y, and Z directions, while the angular velocity data after Kalman filtering has less fluctuation. After filtering, the peak-to-valley difference in the angular velocity in the X direction is reduced from the original 0.25°/s to the filtered 0.17°/s, ensuring more accurate angle measurement. At 0ms, the original angular velocity Z is 5.50°/s, and after Kalman filtering it is 5.52°/s. The magnetometer data also shows small fluctuations. After filtering, the fluctuations of the three-axis magnetic field are further reduced. The peak-to-valley difference in the X direction drops from the original 0.8μT to 0.6μT after filtering, providing more stable environmental data. At 0ms, the original magnetic field X is 45.0μT, and after Kalman filtering, it is 45.2μT. The data optimized by Kalman filtering, combined with the analysis and processing of the CNN model, can effectively improve the accuracy of motion tracking and evaluation, and provide athletes with more accurate real-time feedback and personalized training suggestions.

4.1.Action scoring effect

To comprehensively evaluate the performance of aerobics athletes, a scoring system based on sensor data is introduced. The score value of each sensor at different time points reflects the quality and stability of the athlete's movements. The comprehensive score combines the average score of these three scores to provide a comprehensive performance evaluation for bodybuilders, helping athletes and coaches to grasp the effect of exercise in real-time. Figure 4 shows the scoring performance at different time points.

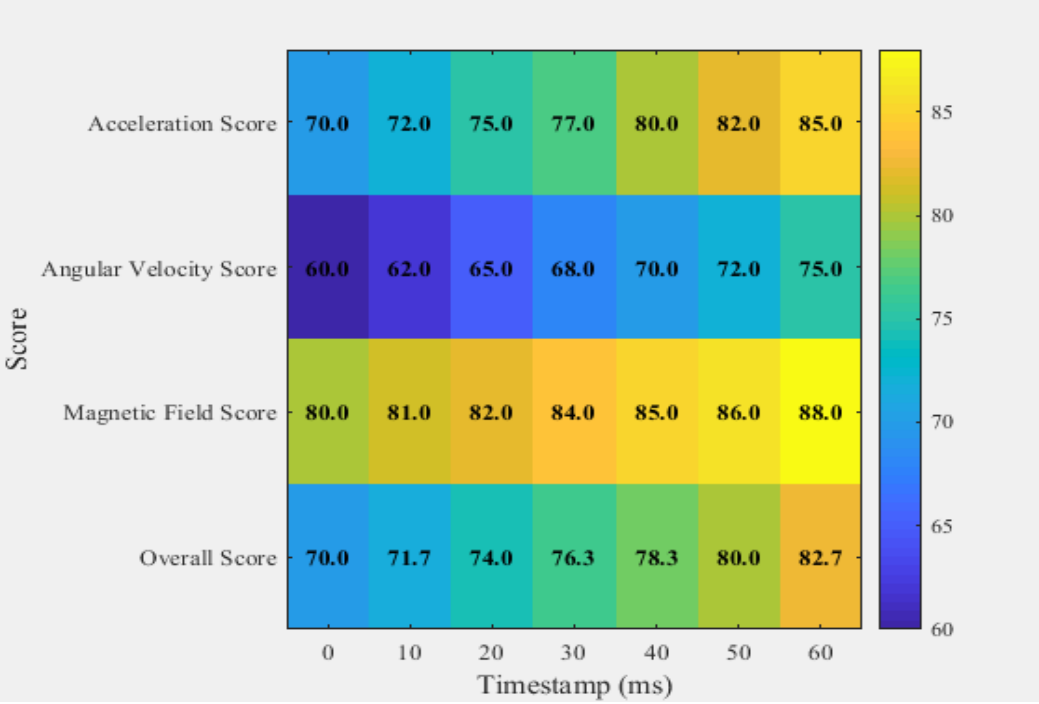


Figure 4: Sensor scores at different time points

In the experiment, wearable sensors and motion tracking technology are combined to fuse and analyze multimodal data such as acceleration, angular velocity, and magnetic field using the CNN model and LSTM to achieve accurate evaluation and optimization of aerobics movements. The data in Figure 4 shows that as time goes by, the athletes' acceleration scores gradually increase from 70 to 85 points; the angular velocity scores increase from 60 to 75; the magnetic field scores also maintain a relatively stable upward trend. The final comprehensive score increases from 70.0 to 82.7. The data changes reflect the gradual optimization of the athletes' performance during training. Figure 4 shows the changes in different scoring dimensions at each time point, helping trainers to accurately monitor and adjust training strategies in real-time. CNN optimizes sensor data by combining Kalman filtering and wavelet transform technology to provide athletes with more accurate performance feedback and promote the improvement of training results.

4.3 Performance of feedback systems in different scenarios

The performance improvement before and after training can reflect the optimization effect of the system. The following Table 5 shows the relationship between feedback delay, optimization suggestion generation time, and athlete performance improvement in different training scenarios.

Table 5: Feedback performance in different training scenarios

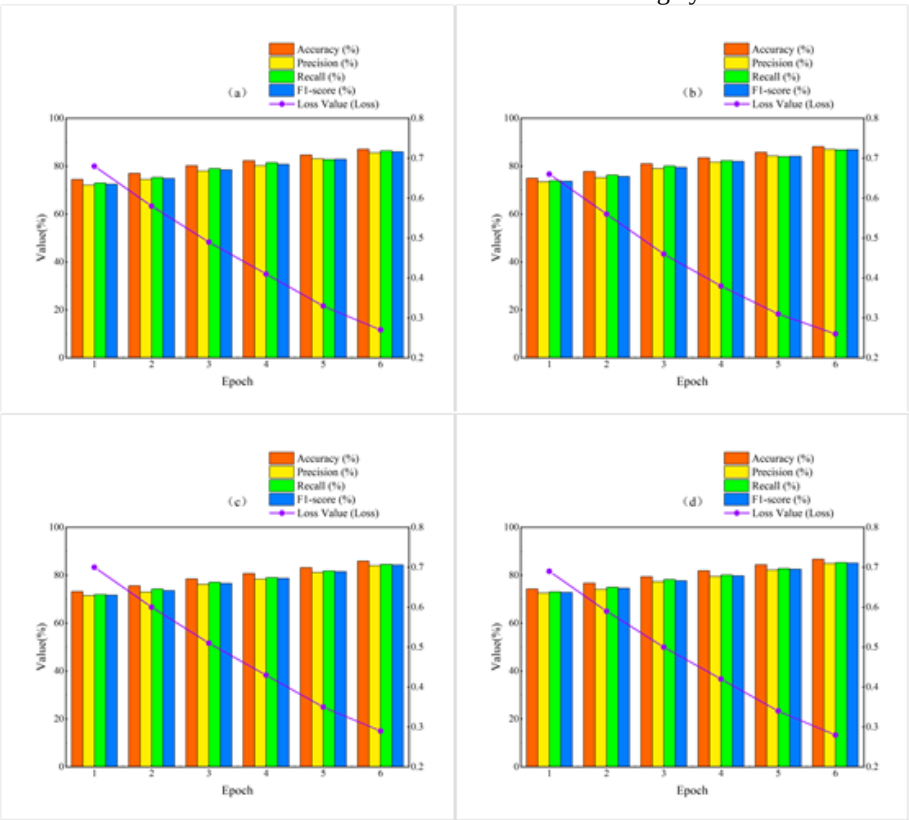
Scenario Type	Average Feedback Delay (ms)	Optimization Suggestion Generation Time (ms)	Pre-training Performance Score (out of 100)	Post-training Performance Score (out of 100)	Improvement Rate (%)
Indoor Low Intensity	150	90	65	80	23.1
Indoor High Intensity	180	95	62	78	25.8
Outdoor Low Intensity	160	85	68	82	20.6
Outdoor High Intensity	200	110	60	75	25

According to the data in Table 5, there are certain differences in feedback delay and optimization suggestion generation time in different training scenarios. In the indoor high-intensity training scenario, the average feedback delay is 180 milliseconds; the optimization suggestion generation time is 95 milliseconds; the action score before training is 62 points. After training, the score increases to 78 points, and the improvement rate reaches 25.8%, which is the highest improvement rate in all scenarios. This shows that under high-intensity training,

despite the longer feedback delay, the system is able to more effectively generate optimization suggestions and improve athlete performance. In outdoor low-intensity training scenarios, the feedback delay is shorter, at 160 milliseconds, but the improvement rate is 20.6%, which is relatively low, indicating that the feedback system in this scenario has room for improvement in the generation and application of optimization suggestions. The efficiency and optimization effect of the feedback system vary in different training scenarios. The response speed of the system and the time to generate suggestions are closely related to the performance improvement of athletes.

4.4 CNN model training effect

To better evaluate and optimize the performance of athletes in different training scenarios, the experiment uses CNN model and LSTM to conduct in-depth analysis of various training data. In the four training scenarios of indoor low intensity, indoor high intensity, outdoor low intensity, and outdoor high intensity, the change of training cycle has an important impact on the accuracy and performance of the model. The changes in the accuracy, precision, recall, F1-score, and loss value of the CNN model in different scenarios can be analyzed to understand the model optimization trend during the training process. Figure 5 shows the changes in key indicators of the model after each training cycle in these scenarios.



(a) Indoor low-intensity training scene; (b) Indoor high-intensity training scene; (c) Outdoor low-intensity training scene; (d) Outdoor high-intensity training scene

Figure 5: CNN effects in different scenes

By analyzing the model effects in different training scenarios through the data in Figure 5, the performance of the CNN model combined with LSTM in various indicators is significantly improved with the increase of training cycles. In the indoor low-intensity training scenario, the accuracy rate increases from 74.5% to 87.1%; the precision, recall rate, and F1-score also increase steadily, and the loss value decreases from 0.68 to 0.27, indicating that the model's ability to fit the data steadily improves with the passage of training time. In indoor high-intensity scenarios, the accuracy rate increases from 75.0% to 88.2%. The improvement in precision and recall rate shows that the model can effectively handle more complex training environments, and the loss value drops to 0.26. The training effect of the model in outdoor low-intensity and high-intensity scenarios also shows a similar trend, with the accuracy rate increased from 73.2% to 85.9% and 74.3% to 86.7%, respectively, and the loss value decreased significantly. Whether it is a low-intensity or high-intensity scenario, the model shows good stability and accuracy in different environments. With the increase of training cycles, the performance of the CNN model in motion scoring and classification has been effectively improved, and the training error can be significantly reduced.

To further verify the effectiveness of the method proposed in this paper in the evaluation of aerobics movements, the CNN-LSTM hybrid model is compared with several representative studies in recent years. The comparative methods include: multimodal fusion method

based on traditional CNN, action recognition method based on LSTM, traditional classification method based on support vector machine (SVM), and temporal modeling method based on Transformer. Comparative experiments are conducted on the same dataset, with evaluation metrics including accuracy and F1 score, as well as action evaluation RMSE (Root Mean Squared Error) value, as shown in Table 6:

Table 6: Comparison of the performance of this method with existing research

Model	Accuracy	F1-score	RMSE
SVM	72.1%	0.703	8.1
CNN	83.2%	0.817	6.0
LSTM	84.6%	0.832	5.8
Transformer	85.4%	0.841	5.6
CNN-LSTM	88.2%	0.867	4.2

From Table 6, it can be seen that this method outperforms existing methods in terms of accuracy and F1 score. Compared with traditional CNN methods, this method has improved accuracy by 5.0%; compared with the LSTM method, it has improved by 3.6%; compared with the Transformer method, it has improved by 2.8%. The RMSE of the method proposed in this paper is 4.2, significantly lower than other comparative methods, indicating that the CNN-LSTM hybrid model proposed in this paper has higher accuracy and stability in predicting action scores. This result represents that the method proposed in this paper can effectively achieve accurate evaluation of athletes' movements and provide objective guidance for scientific training.

5 Experimental discussion

By combining wearable sensors with motion tracking technology and using CNNs for multimodal data fusion and analysis, this study has achieved significant experimental results in motion evaluation and optimization. Kalman filtering significantly improves the stability and accuracy of sensor data in the application of noise suppression, and effectively reduces the impact of environmental interference on data quality. With the increase of training cycles, the CNN model combined with LSTM continues to improve in terms of accuracy, precision, recall rate, and F1-score in action scoring and classification, showing good fitting ability. In high-intensity training scenarios, despite relatively long feedback delays, the system can still effectively generate optimization suggestions, and the athletes' performance has been significantly improved.

However, the experimental results also reveal that there are still some potential challenges and problems in the application and development of the system. The performance of the model has been significantly improved in different training scenarios, but the feedback system has a long delay and optimization suggestion generation time in high-intensity training scenarios, which affects the system's real-time response capability. Kalman filtering and other data optimization techniques effectively reduce noise, but in complex or extreme training environments, external interference still poses a risk of affecting the accuracy of sensor data and causing certain errors in training feedback. With the diversification of training scenarios and the increase in environmental complexity, how to further improve the model's ability to classify complex actions and its ability to comprehensively process multimodal data remains an issue to be resolved. Future research can focus on optimizing sensor data collection and processing technology, strengthening data synchronization and fusion algorithms, and further improving the stability and adaptability of CNN models in different environments. With the advancement of technology, combined with more sensors and analysis of training scenarios, it can be possible to provide athletes with more detailed and comprehensive personalized training programs, promoting the intelligent and precise development of sports training.

6 Conclusions

This paper proposes a fitness exercise action evaluation method that integrates wearable sensors and motion tracking systems, and combines CNN and LSTM models to fuse and analyze multimodal data. The experimental results show that the method exhibits excellent performance in action classification tasks: in indoor low-intensity training scenarios, the accuracy increases from 74.5% to 87.1%; in high-intensity training scenarios, the accuracy increases from 75.0% to 88.2%. By introducing Kalman filtering, wavelet transform, and dynamic weighting fusion strategy, the stability of sensor data and the generalization ability of the model have been effectively improved. This paper not only provides high-precision motion evaluation and real-time feedback for aerobics athletes, but also provides a transferable technical framework for other high-intensity, multimodal motion sports projects. In the future, the system can be further expanded to remote training platforms, intelligent wearable devices, virtual coaching systems, and other application scenarios, promoting the deep integration and widespread application of artificial intelligence technology in the fields of sports training and health management. In the future, lightweight CNN structures such as MobileNet and TinyML, deployment of models, on wearable devices, and heterogeneous computing acceleration can be used to further shorten feedback latency and improve system response speed.

Authorship contribution statement

Danhua Tan: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Data availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

References

[1] M. E. M. Simbolon, D. K. A. Firdausi, I. Dwisaputra, A. Rusdiana, C. Pebriandani, and R. Prayoga, "Utilization of Sensor technology as a Sport Technology Innovation in Athlete Performance Measurement," *Indonesian Journal of Electronics and Instrumentation Systems (IJEIS)*, 13(2): 147–158, 2023. https://doi.org/10.22146/ijeis.89581

[2] Z. Mei, "3D images analysis of sports technical features and sports training methods based on artificial intelligence," *J Test Eval*, 51(1): 189–200, 2023. https://doi.org/10.1520/JTE20210469

[3] S. A. Kovalchik, "Player tracking data in sports," *Annu Rev Stat Appl*, 10(1): 677–697, 2023. https://doi.org/10.1146/annurev-statistics-033021-110117

[4] L. Yang, O. Amin, and B. Shihada, "Intelligent wearable systems: Opportunities and challenges in health and sports," *ACM Comput Surv*, 56(7):1–42, 2024. https://doi.org/10.1145/3648469

[5] W. Li, "Application of IoT-enabled computing technology for designing sports technical action characteristic model," *Soft comput*, 27(17): 12807–12824, 2023. https://doi.org/10.1007/s00500-023-08966-4

[6] Y. Fang, "Utilizing Wearable Technology to Enhance Training and Performance Monitoring in Indonesian Badminton Players," *Studies in Sports Science and Physical Education*, 2(1): 11–23, 2024. DOI:10.1186/s40561-023-00247-9

[7] J. Corban *et al.*, "Using an affordable motion capture system to evaluate the prognostic value of drop vertical jump parameters for noncontact ACL injury," *Am J Sports Med*, 51(4):1059–1066, 2023. https://doi.org/10.1177/03635465231151686

[8] C. J. Rigozzi, G. A. Vio, and P. Poronnik, "Application of wearable technologies for player motion analysis in racket sports: A systematic review," *Int J Sports Sci Coach*, 18(6): 2321–2346, 2023. https://doi.org/10.1177/17479541221138015

[9] Y. Zhang, "Design of Wireless Motion Sensor Nodes based on the Kalman Filter Algorithm," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, 16(3): 248–255, 2023. https://doi.org/10.2174/2352096515666220908152036

[10] S. Akan and S. Varlı, "Use of deep learning in soccer videos analysis: survey," *Multimed Syst*, 29(3): 897–915, 2023. https://doi.org/10.1007/s00530-022-01027-0

[11] D. Gholamiangonabadi and K. Grolinger, "Personalized models for human activity recognition with wearable sensors: deep neural networks and signal processing," *Applied Intelligence*, 53(5): 6041–6061, 2023. https://doi.org/10.1007/s10489-022-03832-6

[12] A. Chakraborty and N. Mukherjee, "A deep-CNN based low-cost, multi-modal sensing system for efficient walking activity identification," *Multimed Tools Appl*, 82(11): 16741–16766, 2023. https://doi.org/10.1007/s11042-022-13990-x

[13] W. Liu, Y. Liu, and R. Bucknall, "Filtering based multi-sensor data fusion algorithm for a reliable unmanned surface vehicle navigation," *Journal of Marine Engineering & Technology*, 22(2): 67–83, 2023. https://doi.org/10.1080/20464177.2022.2031558

[14] L. Zhang and H. Dai, "Motion trajectory tracking of athletes with improved depth information-based KCF tracking method," *Multimed Tools Appl*, 82(17): 26481–26493, 2023. https://doi.org/10.1007/s11042-023-14929-6

[15] P. Hao and K. Qian, "The Integration of Personalized Training Program Design and Information Technology for Athletes," *Scalable Computing: Practice and Experience*, 25(5): 4351–4359, 2024. https://doi.org/10.12694/scpe.v25i5.3083

[16] V. Deepak, D. K. Anguraj, and S. S. Mantha, "An efficient recommendation system for athletic performance optimization by enriched grey wolf optimization," *Pers Ubiquitous Comput*, 27(3): 1015–1026, 2023. https://doi.org/10.1007/s00779-022-01680-2

[17] J. K. Urbanek *et al.*, "Free-living gait cadence measured by wearable accelerometer: a promising alternative to traditional measures of mobility for assessing fall risk," *The Journals of Gerontology: Series A*, 78(5): 802–810, 2023. https://doi.org/10.1093/gerona/glac013

[18] A. Hussain, S. Ali, M.-I. Joo, and H.-C. Kim, "A deep learning approach for detecting and classifying cat activity to monitor and improve cat's well-being using accelerometer, gyroscope, and magnetometer," *IEEE Sens J*, 24(2): 1996–2008, 2023.

[19] A. Spilz and M. Munz, "Synchronisation of wearable inertial measurement units based on magnetometer data," *Biomedical Engineering/Biomedizinische Technik*, 68(3): 263–273, 2023. https://doi.org/10.1515/bmt-2021-0329

[20] A. Liu, R. P. Mahapatra, and A. V. R. Mayuri, "Hybrid design for sports data visualization using AI and big data analytics," *Complex & Intelligent Systems*, 9(3): 2969–2980, 2023. https://doi.org/10.1007/s40747-021-00557-w

[21] C.-T. Lin, Y. Wang, S.-F. Chen, K.-C. Huang, and L.-D. Liao, "Design and verification of a wearable wireless 64-channel high-resolution EEG acquisition system with wi-fi transmission," *Med Biol Eng Comput*, 61(11): 3003–3019, 2023. https://doi.org/10.1007/s11517-023-02879-y

[22] X. Shi and H. Zou, "Data Collection and Analysis based on Sensor Technology in Sports Training," *Scalable Computing: Practice and Experience*, 25(5): 4399–4406, 2024. https://doi.org/10.12694/scpe.v25i5.3200

[23] M. Khodarahmi and V. Maihami, "A review on Kalman filter models," *Archives of Computational Methods in Engineering*, 30(1): 727–747, 2023. https://doi.org/10.1007/s11831-022-09815-7

[24] M. Azhar, S. Ullah, M. Raees, K. U. Rahman, and I. U. Rehman, "A real-time multi view gait-based automatic gender classification system using kinect sensor," *Multimed Tools Appl*, 82(8): 11993–12016, 2023. https://doi.org/10.1007/s11042-022-13704-3

[25] L. Lv, J. Yang, F. Gu, J. Fan, Q. Zhu, and X. Liu, "Validity and reliability of a depth camera–based quantitative measurement for joint motion of the hand," *J Hand Surg Glob Online*, 5(1): 39–47, 2023. https://doi.org/10.1016/j.jhsg.2022.08.011

[26] Y. Wu, Z. Sun, G. Ran, and L. Xue, "Intermittent control for fixed-time synchronization of coupled networks," *IEEE/CAA Journal of Automatica Sinica*, 10(6): 1488–1490, 2023. DOI: 10.1109/JAS.2023.123363

[27] Hrovatin, N. "Enabling Decentralized Privacy Preserving Data Processing in Sensor Networks," Informatica (03505596), 48(1): 141-142, 2024. https://doi.org/:10.31449/inf.v48i1.5739.

[28] Thi H N, Duc C V, Duc C T, HH Minh, SN Van, LV Quan. "Memetic Algorithm for Maximizing K-coverage and K-Connectivity in Wireless Sensor Network," Informatica, (03505596), 49(1): 1-7, 2025. https://doi.org/:10.31449/inf.v49i1.6750.

[29] A. Halidou, Y. Mohamadou, A. A. A. Ari, and E. J. G. Zacko, "Review of wavelet denoising algorithms," *Multimed Tools Appl*, 82(27): 41539–41569, 2023. https://doi.org/10.1007/s11042-023-15127-0

[30] M. Kang, C. L. Bentley, J. T. Mefford, W. C. Chueh, and P. R. Unwin, "Multiscale Analysis of Electrocatalytic Particle Activities: Linking Nanoscale Measurements and Ensemble Behavior," *ACS Nano*, 17(21): 21493–21505, 2023. https://doi.org/10.1021/acsnano.3c06335

[31] J. Sun, H. Zhang, X. Ma, R. Wang, H. Sima, and J. Wang, "Spectral–Spatial Adaptive Weighted Fusion and Residual Dense Network for hyperspectral image classification," *The Egyptian Journal of Remote Sensing and Space Sciences*, 28(1): 21–33, 2025. https://doi.org/10.1016/j.ejrs.2024.11.001

[32] J. P. Bharadiya, "A tutorial on principal component analysis for dimensionality reduction in machine learning," *Int J Innov Sci Res Technol*, 8(5): 2028–2032, 2023. DOI:10.5281/zenodo.8002436

[33] F. Bizzarri, D. Del Giudice, S. Grillo, D. Linaro, A. Brambilla, and F. Milano, "Inertia estimation through covariance matrix," *IEEE Transactions on Power Systems*, 39(1): 947–956, 2023. DOI: 10.1109/TPWRS.2023.3236059

[34] Y. He, C.-K. Zhang, H.-B. Zeng, and M. Wu, "Additional functions of variable-augmented-based free-weighting matrices and application to systems with time-varying delay," *Int J Syst Sci*, 54(5): 991–1003, 2023. https://doi.org/10.1080/00207721.2022.2157198

[35] Singh S, Sehgal V K. "Deep Learning-Based CNN Multi-Modal Camera Model Identification for Video Source Identification," Informatica: An International Journal of Computing and Informatics, 47(3): 417-430, 2023. https://doi.org/10.31449/inf.v47i3.4392

[36] T. Sharma, N. K. Verma, and S. Masood, "Mixed fuzzy pooling in convolutional neural networks for image classification," *Multimed Tools Appl*, 82(6): 8405–8421, 2023. https://doi.org/10.1007/s11042-022-13553-0

[37] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput Appl*, 35(23): 17095–17112, 2023. https://doi.org/10.1007/s00521-023-08568-z

[38] Z. Mei *et al.*, "Automatic loss function search for adversarial unsupervised domain adaptation," *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10): 5868–5881, 2023. DOI: 10.1109/TCSVT.2023.3260246