SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(4), 1396.
https://doi.org/10.62617/mcb1396

Article

A health monitoring and exercise optimization method based on fusion of action recognition and physiological signals—A study on students at the Zhejiang police college

Ganbin Xu[1], Weihong Li[2,*], Xia Li[3]

[1] Zhejiang Police College, Hangzhou 310000, China
[2] Department of Scientific Research and Local Cooperation, Lishui Vocational and Technical College, Lishui 323000, China
[3] College of Marxism, Zhejiang Oriental Vocational and Technical College, Wenzhou 325000, China
*** Corresponding author:** Weihong Li, 971668916@163.com

CITATION

Xu G, Li W, Li X. A health monitoring and exercise optimization method based on fusion of action recognition and physiological signals—A study on students at the Zhejiang police college. Molecular & Cellular Biomechanics. 2025; 22(4): 1396.
https://doi.org/10.62617/mcb1396

ARTICLE INFO

Received: 22 January 2025
Accepted: 3 March 2025
Available online: 17 March 2025

Abstract: With the growing emphasis on health management and the widespread adoption of smart devices, the demand for motion recognition and health anomaly monitoring has become increasingly pressing. In particular, for students at the Zhejiang Police College, maintaining good physical fitness and health is critical because they often face high-intensity work pressure and unpredictable work environments. Fundamentally, health monitoring is a pattern recognition challenge, wherein the objective is to extract critical features from multimodal data to detect abnormal behaviors or health conditions, thereby enabling real-time surveillance of individual health statuses. Against this backdrop, this paper introduces a novel health monitoring framework (PSCLM), which integrates convolutional neural networks (CNN), long short-term memory networks (LSTM), and modal decomposition techniques, aiming to enhance the precision and resilience of action recognition and health anomaly detection. Initially, the proposed framework analyzes inertial sensor data and constructs a convolutional long short-term memory (CLM) model utilizing CNN and LSTM to facilitate movement type recognition. Subsequently, given the significance of heart rate as a vital indicator during physical activity, heart rate and electrocardiogram (ECG) signals, along with their variational mode decomposition (VMD) decomposition features, are incorporated with movement recognition features to achieve multidimensional and hierarchical fusion of physiological signals during exercise. Finally, leveraging these fused features, the framework achieves comprehensive monitoring of exercise-induced health states. Experimental evaluations reveal that the PSCLM framework attains an average accuracy exceeding 0.9 for action recognition on publicly available datasets, while its accuracy for abnormal state detection reaches 0.95, outperforming traditional approaches and single deep learning models significantly. This research not only provides an innovative technological approach for sports health management and anomaly detection but also establishes a critical foundation for advancing next-generation intelligent health monitoring systems.

Keywords: motion recognition; signal fusion; health monitoring; deep learning

1. Introduction

In recent years, the heightened public awareness of health has catalyzed the rapid evolution of health monitoring technologies. The Zhejiang Police College trains future law enforcement and public safety professionals who require not only exceptional physical fitness but also strong psychological resilience and rapid response capabilities. To meet the demands of their future careers, students at China

People's Police University (CPPU) must undergo rigorous physical training. Transitioning from traditional periodic physical examinations to contemporary real-time, continuous monitoring systems, health monitoring has undergone transformative advancements. Modern systems are grounded in the precise acquisition and analysis of diverse physiological and behavioral signals, including heart rate (HR), electrocardiographic signals (ECG), body temperature, blood oxygen saturation, and human movement patterns [1].

Among these, HR and ECG signals are pivotal to health monitoring. HR reflects the frequency and regularity of cardiac activity, serving as a fundamental indicator of cardiovascular health, while ECG provides detailed waveforms of the heart's electrical activity, enabling accurate detection of conditions such as arrhythmia and myocardial ischemia. Together, these signals offer a comprehensive assessment of cardiac function and overall health, forming a critical foundation for early disease warning, diagnosis, and treatment. Moreover, they play an essential role in sports health management and stress evaluation. Real-time, precise monitoring of HR and ECG signals constitutes a cornerstone of personalized health management. These signals encapsulate an individual's health status, delivering vital insights for disease prevention, chronic disease management, and timely health interventions. With ongoing technological advancements, portable sensors, particularly wearable devices, have emerged as the nucleus of modern health monitoring. These innovations not only enhance the personalization of monitoring but also broaden the scope of health data acquisition and analysis, paving the way for intelligent health management systems [2].

Wearable devices are increasingly prevalent in the field of health monitoring, leveraging embedded sensors to continuously capture human movement and physiological information. The sheer volume and complexity of data generated by these devices present significant challenges for traditional data processing methods, which often struggle to fully extract and utilize the latent information. To address this, machine learning and deep learning techniques have been extensively adopted in recent years for health signal analysis [3]. Among these, CNN is widely employed to extract spatial features, while LSTM excels in modeling temporal dependencies within time-series data. Additionally, methods such as autoencoders and variational autoencoders (VAEs) are frequently utilized for feature dimensionality reduction and anomaly detection. These advanced techniques have substantially enhanced the extraction of critical insights from multimodal health data, significantly improving the efficiency and accuracy of health monitoring systems [4]. Beyond inertial data, wearable technology advancements now enable the real-time acquisition of comprehensive physiological indicators such as ECG signals and heart rate during exercise. Further feature extraction using methods like empirical mode decomposition (EMD), variational mode decomposition (VMD), empirical wavelet transform (EWT), and wavelet transform plays an indispensable role in efficient health monitoring and precise anomaly detection, driving the development of next-generation health management solutions.

Traditional health monitoring methods often face challenges in effectively integrating multimodal physiological data and providing real-time, accurate assessments. Many conventional systems rely on rule-based models or simple statistical approaches, which struggle to capture complex relationships between physiological signals, leading to potential inaccuracies in health status evaluations. Furthermore, existing machine learning-based methods often lack efficient real-time processing capabilities, limiting their ability to adapt to dynamic physiological changes. To address the health monitoring challenges related to motion optimization, this paper investigates physiological signals collected by wearable devices, completing research on motion recognition and early warning of health anomalies. A multimodal fusion approach is proposed, integrating inertial sensor data with physiological signals to enhance the accuracy of motion recognition and health anomaly monitoring. The primary contributions of this study are as follows:

1) A motion recognition framework is developed, leveraging CNN and LSTM to achieve high-precision classification of motion types, tailored to the needs of multimodal motion recognition and health anomaly detection.
2) Building on the CLM framework, a system is established for abnormal health status monitoring, incorporating physiological data such as heart rate, ECG, and multi-level movement features, enabling the detection of health anomalies in physically active individuals.
3) Experimental results on public datasets demonstrate that the integration of physiological signals significantly enhances the action recognition accuracy of the CLM framework. The fusion of multi-level recognition results further improves the model’s performance, surpassing that of traditional methods.

The remainder of this paper is structured as follows: Section 2 reviews related works on motion recognition and health monitoring. Section 3 presents the proposed frameworks, CLM and PSCLM, detailing their design and implementation. Section 4 provides a comprehensive description of the experimental setup and results. Finally, the paper concludes with a summary of findings and future research directions.

2. Related works

2.1. Sensor-based action recognition monitoring

The research and application of real-time physiological health monitoring techniques can be broadly categorized into motion recognition and measurement, abnormality detection and prediction, and clinical applications. Among these, inertial sensors are the most commonly employed in motion monitoring research. For instance, Adlian Jefiza et al. developed a fall detection device for the elderly using the MPU6050 motion sensor, which integrates a three-axis accelerometer and gyroscope to capture both rapid and slow movements. This device employs a backpropagation algorithm to recognize multiple postures, including leaning forward, sideways, and backward, sitting, sleeping, crouching, climbing stairs, and praying, with the ability to quickly detect falls during rest [5]. Shun Ishii et al. proposed a real-time segmentation, classification, and measurement algorithm for sports using IMU sensors. This algorithm applies time windows and synthetic acceleration peaks to segment motion data streams, classifies motion types using template signals and the dynamic time warping (DTW) algorithm, and recognizes five types of movements—walking, running, jumping, push-ups, and sit-ups—while also counting the repetitions of each action [6]. Dawar et al. introduced a CNN-based sensor fusion system that detects and recognizes actions of interest within a continuous action stream, employing decision-level fusion to achieve action recognition [7]. Similarly, Liu et al. enhanced gesture recognition accuracy by integrating inertial and visual sensor data within a Hidden Markov Model framework [8]. Li et al. demonstrated the benefits of fusing feature information extracted from different sensor data, addressing the performance limitations of single sensors by combining heterogeneous sensor data to improve overall system performance [9]. Ehatisham et al. proposed a feature-level fusion method for human action recognition, utilizing visual and inertial data from two distinct perception modalities. Their supervised machine learning approach effectively fuses features extracted from each modality to achieve precise action recognition [10]. Additionally, Radu et al. employed four different deep neural network variants to interpret user activity and context detection using data from multiple sensors [11].

From the above research, it can be seen that with the continuous enrichment of data types, the field of action recognition is no longer limited to a single video stream or inertial sensor data, but is gradually developing towards multimodal data fusion. The combination of different data sources provides more comprehensive information for action recognition, but how to fully utilize this data to improve recognition accuracy remains a key challenge in monitoring tasks. This study suggests that optimizing the selection of sensors and improving information utilization efficiency through multi-level data fusion can more effectively leverage the advantages of machine learning and deep learning. Therefore, optimizing signal features is an important way to improve detection performance.

2.2. Motion monitoring analysis based on physiological signal analysis

For physiological signal monitoring studies, assessing several fundamental vital signs—pulse, respiration, blood pressure, temperature, and oxygen saturation—is critical for evaluating an individual’s physical health. These parameters are key indicators of a patient’s condition and its severity. Reza et al. proposed a clinical decision support system embedded in a wrist-worn device to predict sleep quality by analyzing changes in physiological signals during deep sleep stages. The device monitored metrics such as movement status, heart rate, body surface temperature, and skin electrical activity, demonstrating its efficacy in helping users improve sleep quality and rest [12]. NHO et al. introduced an innovative solution for fall detection [13]. Meanwhile, Yunsik et al. developed a heart rate recovery evaluation system capable of real-time motion state classification. Their system combined acceleration and heart rate data, extracted features for analysis, and utilized peak acceleration and peak angular tilt for motion classification, inspiring advancements in this domain [14]. Xiao et al. designed a heart rate monitoring bracelet for exercise scenarios using technologies like Bluetooth, IoT, and ZigBee. The device transmitted motion and heart rate data to smartphones and PCs for storage and analysis, enabling real-time heart rate health monitoring and abnormal data alerts [15]. Dimitris et al. proposed a self-supervised representation learning method, using heart rate signals as supervised inputs to learn generalizable features from large-scale data, uncovering relationships between exercise signals and heart rate [16]. Woosuk et al. conducted a study on continuous physiological health monitoring using a smart bracelet to track daily activities and changes in exercise, heart rate signals, and BMI over an extended period. Regression analysis of obesity-related indices highlighted its utility in personal health management [17]. Other notable studies include Zhong Ke et al., who utilized ECG signals to detect driving fatigue [18], and Mihaela et al., who employed ECG to identify hypoglycemic events [19].

These studies underscore the vast potential of physiological monitoring sensors in various applications, from health management to clinical and lifestyle interventions, showcasing their transformative impact on modern healthcare systems. It is easy to see from the above study that the research on human health and physiological state monitoring using multi-source and multi-modal information has a wider prospect. Therefore, the fusion of physiological signals on the basis of traditional action recognition can lead to more accurate and diversified task types. In this paper, we propose to use the fusion of action signals and physiological signals to observe the effect of different physiological signals on motion recognition, which is of great significance for the research of health monitoring and exercise refinement.

3. Methodology

3.1. CNN-LSTM (CLM) based action recognition

3.1.1. Inertial data and feature processing

Building on the existing characteristics of action recognition research, we first utilize inertial data to perform action recognition. The results of this recognition are then input into a corresponding physiological state monitoring model for further analysis, enabling multi-level data fusion to comprehensively monitor and analyze the movement states of the target [20]. To enhance data quality, we preprocess the inertial data by applying a Butterworth filter to the acceleration signals. This filter effectively eliminates high-frequency or low-frequency noise, preserving the primary frequency components of the signal to ensure accurate and reliable data for subsequent processing.

For the input signal $x[n]$, the output of the filter $y[n]$ is calculated by the difference equation:

$$y[n] = \sum_{i=0}^{N_b} b[i]\, x[n-i] - \sum_{j=1}^{N_a} a[j]\, y[n-j] \tag{1}$$

where $b[i]$ and $a[j]$ are the numerator and denominator coefficients of the filter. The Butterworth data can be used to effectively filter the noise collected by the IMU to ensure the stability of the signal, for a typical walking process, the acceleration in the direction of the forward sides and its corresponding filtered data are shown in **Figure 1**:

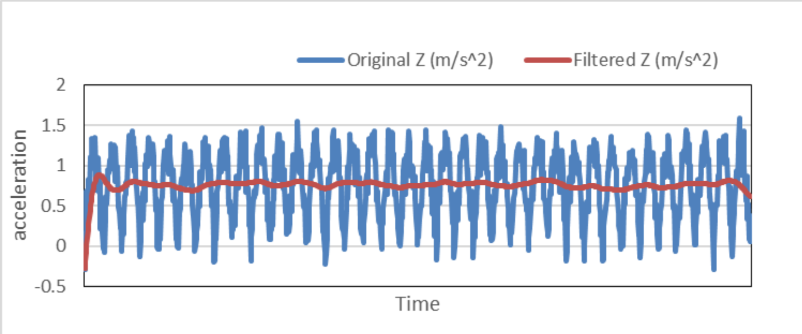

Figure 1. The data processed by the butterworth filter.

Following the filtering process to eliminate noise, we employ a CNN to extract the key features of the signals. Furthermore, based on existing research, it has been observed that the intensity of motion during physical activity—represented by the magnitude of acceleration in various directions—can be more effectively analyzed using the Euclidean norm (two-norm). The computation for this process is expressed as follows:

$$Intensity = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{2}$$

Therefore, feature extraction and corresponding action classification studies are accomplished with the help of deep neural networks. This study utilizes acceleration and angular velocity signals as input data for action recognition. First, data normalization and noise filtering were applied to enhance signal stability, followed by the extraction of intensity-based features to capture variations in movement magnitude and dynamics, providing crucial support for action recognition. To effectively leverage this information, we constructed a CNN-LSTM hybrid model, where the CNN component extracts local temporal features, and the LSTM component further learns long-term dependencies in the time series to improve recognition accuracy. Specifically, the CNN consists of two one-dimensional convolutional layers with 64 and 128 filters, respectively, using a kernel size of 3 and ReLU activation. A max pooling layer with a pooling window size of 2 follows each convolution to reduce data dimensionality. Subsequently, two LSTM layers with 128 and 64 hidden units, respectively, capture temporal dependencies in movement patterns. After flattening, the data is passed through a fully connected layer with 128 neurons and finally processed through a Softmax layer to output the classification results. During model training, we used the Adam optimizer with a learning rate of 0.001, a batch size of 64, and 50 training epochs. Cross-entropy loss was employed to optimize model performance.

3.1.2. The motion recognition model

After completing data preprocessing and feature determination, we proceeded to construct the action recognition model. The model employs a CNN to extract spatial features and capture local patterns. Subsequently, an LSTM is utilized to handle long-term dependencies in the time series, model the dynamic characteristics of actions, and accomplish the classification task.

The CNN extracts features from the local time series through one-dimensional convolution, with the convolution operation defined as follows:

$$H_t = \sigma(W_c \times X_t + b_c) \tag{3}$$

where, W_c is the kernel weights and b_c is the bias. After completing the convolutional extraction of features, further sampling, i.e., a pooling operation, is performed to downsample the local features to reduce the data dimensions and retain the important features.

$$H_p = \text{pooling}(H_t) \tag{4}$$

After completing the corresponding pooling, the corresponding output feature matrix can be obtained as shown in Equation (5):

$$F = \{f_1, f_2, \ldots, f_M\} \tag{5}$$

The obtained M-dimensional features are fed into the LSTM network for time series feature learning, for which feature extraction and updating are mainly performed through forgetting gates, input gates and output gates.

$$f_t = \sigma\left(W_f h_{t-1} + U_f f_{t-1} + b_f\right) \tag{6}$$

$$i_t = \sigma(W_i h_{t-1} + U_i f_t + b_i) \tag{7}$$

$$o_t = \sigma(W_o h_{t-1} + U_o f_t + b_o) \tag{8}$$

where f_t is the forgetting gate, it is the input gate, the ot dimensional output gate, through the above three gates to control the input and output of the model and thus carry out the extraction of long time series features. On this basis, we utilize the softmax function to complete the corresponding action classification. The whole flow of the action recognition module is shown in **Figure 2**:

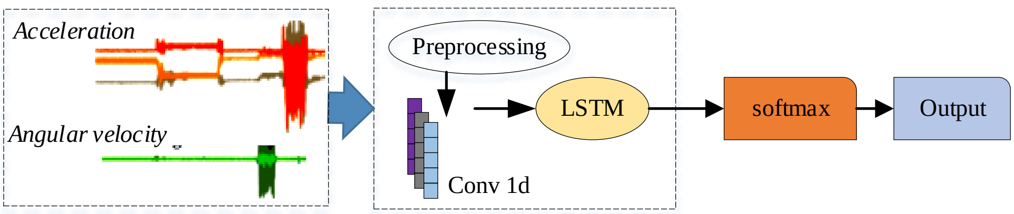

Figure 2. The framework CLM for the motion recognition.

The CNN-LSTM motion recognition framework proposed begins by filtering acceleration and angular velocity inertial data to remove noise. The processed data is then input into a 1D-CNN network for feature enhancement, extracting spatial and local temporal patterns. Subsequently, the enhanced features are passed through a two-layer LSTM network, which models long-term dependencies and dynamic characteristics of motion. Finally, the framework completes the corresponding action classification and recognition, providing decision-layer features for subsequent health monitoring tasks. This approach significantly enhances monitoring stability and reliability.

3.2. A health monitoring model that incorporates heart rate ECG and exercise information (PSCLM)

After constructing the motion action recognition module, the next step involves detecting the physical condition of the subject based on their physiological signals. The PSCLM framework for health monitoring, illustrated in **Figure 3**, integrates the results of CLM action recognition at the decision-making level. This framework facilitates the fusion and analysis of data by incorporating basic ECG and heart rate signals, enabling a comprehensive assessment of the subject's health status during activity.

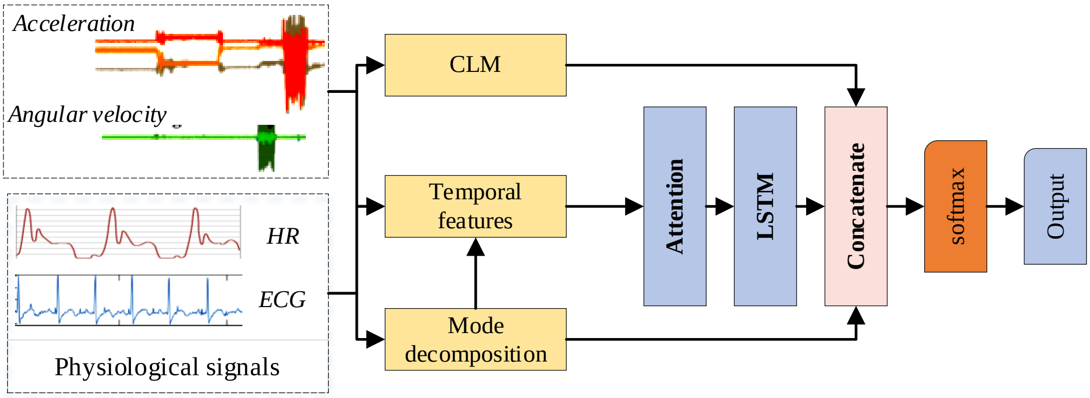

Figure 3. The framework for the health monitoring PSCLM.

Building on the CLM framework described in Section 3.1, ECG and heart rate physiological signals were incorporated into the health monitoring process, as these are crucial for risk detection in subjects [21]. By reviewing relevant studies and analyzing the data types available in current public datasets, ECG and heart rate signals were selected for further analysis. For processing the ECG signals, VMD was applied to perform eigenmode decomposition, effectively extracting key features from the signals for subsequent monitoring and analysis.

VMD decomposes the signal into K modes, and each mode $u_k(t)$ corresponds to a center frequency ω_k. Through iterative optimization, the VMD minimizes the bandwidth of each mode to achieve signal decomposition [22]. The goal of the VMD is to reconstruct the original signal by optimizing the sum of all modes by minimizing the bandwidth of each mode:

$$\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times \left(u_k(t) e^{-j\omega_k t} \right) \right] \right\|_2^2 \tag{9}$$

where: $u_k(t)$ is the modal signal. ω_k indicates the center frequency. ∂_t It is used to calculate the bandwidth. $\delta(t)$ Then denotes the unit pulse signal No. The VMD first initializes the signal K the signal of each mode $u_k(t)$ and the center frequency ω_k while setting the Lagrange multiplier $\lambda(t)$ and the constraint parameter α. The obtained constructed Lagrangian function is shown in Equation (10):

$$\mathrm{L}\left(\{u_k\},\{\omega_k\},\lambda\right)=\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)\times\left(u_k(t)e^{-j\omega_k t}\right)\right]\right\|_2^2+\left\|f(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2 \tag{10}$$

where $f(t)$ is the original signal. For each mode $u_k(t)$:

$$u_k(t)=\arg\min_{u_k}\mathrm{L}\left(\{u_k\},\{\omega_k\},\lambda\right) \tag{11}$$

Update center frequency ω_k:

$$\omega_k=\frac{\int_{-\infty}^{\infty}|t|\left|u_k(t)\right|^2 dt}{\int_{-\infty}^{\infty}\left|u_k(t)\right|^2 dt} \tag{12}$$

Update Lagrange multipliers λ:

$$\lambda(t)=\lambda(t)+\alpha\left[f(t)-\sum_{k=1}^{K}u_k(t)\right] \tag{13}$$

The iteration is stopped when the update amplitude of $u_k(t)$ or ω_k is smaller than the preset threshold. The loss function used is the cross-entropy loss function, and its calculation process can be expressed by Equation (14):

$$J=-\sum_{i=1}^{K}-\beta_i y_i\log\left(p_i\right) \tag{14}$$

where y represents the corresponding white label, p is the network output while K represents the corresponding number of categories and β is the adjustable hyperparameter. After performing the VMD decomposition of the ECG signal, features from each decomposed component are extracted. Additionally, statistical features are derived from these components and fused with the statistical features of the inertial data. This fused data is then fed into an attention module to enhance feature relevance, followed by processing through an LSTM network for further feature extraction. Finally, the motion recognition features obtained from the CLM framework are integrated with the original modal decomposition features to complete the comprehensive motion state monitoring process.

4. Experiment result and analysis

4.1. Dataset and experiment setup

The PAMAP2 dataset is a comprehensive motion monitoring dataset designed for recognizing a wide array of daily activities. It includes physiological signals such as motion data and heart rate, collected using three IMUs placed on the arms, chest, and feet, along with heart rate monitors. The dataset provides triaxial acceleration, gyroscope, magnetometer, and HR signals, sampled at frequencies of 100 Hz and 9 Hz, respectively. Spanning 18 activity labels, including static activities (e.g., sitting, standing) and dynamic activities (e.g., walking, running, biking) [23], PAMAP2 is well-suited for studying multimodal data fusion, such as integrating motion signals and heart rate information to evaluate exercise intensity and health status. Its versatility makes it widely applicable in areas such as exercise optimization, movement classification, and health monitoring.

Similarly, the MHEALTH dataset is a multimodal action recognition dataset developed for studying daily activities and health monitoring. It comprises data collected through wearable devices, covering multiple physiological and motion signals, including triaxial acceleration, triaxial gyroscope, triaxial magnetometer, and ECG. Recorded at a sampling frequency of 50 Hz, the dataset includes 12 common movement labels, such as walking, running, ascending and descending stairs, sitting, standing, and lying down [24]. A distinctive feature of this dataset is its combination of action signals with physiological signals like ECG, making it particularly valuable for exploring multimodal fusion applications in action classification and health state evaluation.

After introducing the dataset, we proceed to monitor the corresponding health status, considering future applications in sports and related requirements. The final output of the PSCLM model is categorized into three classes: High-intensity, medium-intensity, and low-intensity exercise. Ten daily activities from the dataset, including walking, running, and ascending or descending stairs, were selected for classification into high- and medium-intensity categories. These labels were used to distinguish the current movement state of the subject. If the model detects smooth movement but the heart rate and ECG signals correspond to high-intensity activity, the state is flagged as abnormal.

The labeling of activities and their intensity classifications were analyzed and validated by medical professionals based on heart rate and ECG signals. To address the issue of limited anomalous data, a portion of the running heart rate data was intentionally relabeled as walking data to artificially increase the volume of health anomaly data. This approach mitigates the risk of model overfitting and ensures better generalization for health state anomaly detection. The dataset used in this study is publicly available and has been anonymized to ensure compliance with data-sharing regulations. No personally identifiable information is included, eliminating privacy concerns related to data collection and usage.

Following the description of the dataset and corresponding data processing, we first evaluated the action recognition methods. The comparison included CNN, LSTM, and classical machine learning approaches such as support vector machines (SVM) and backpropagation neural networks (BPNN). For anomaly monitoring, the proposed framework was compared against the methods introduced by Lin et al. [21] and Ali et al. [25], with detailed analysis provided accordingly.

Additionally, ablation experiments were conducted to assess the contributions of specific components within the framework. These included the PSCLM-V variant, which excludes the VMD decomposition information, and the PSCLM-R framework, which omits the fused decision features. The results from these ablation experiments provided further insight into the effectiveness of the individual components of the PSCLM model. Considering the multi-source and multi-modal nature of the data and the corresponding amount of data, we build the experimental environment shown in **Table 1**:

Table 1. The Experiment environment information.

Item	Specification
CPU	I7-14400F
GPUs	RTX 4090
Language	Python 3.5.1
Framework	Pytorch

In terms of evaluation metrics for the model, we use accuracy in the motion recognition framework to analyze the recognition accuracy of the CLM model. This is because for the initial judgment and analysis of the model, only high accuracy can have certain advantages in the subsequent fine-tuning process. Therefore, in the subsequent health monitoring, we identify abnormal movements to ensure the health monitoring of athletes.

4.2. Motion recognition result and analysis

After constructing and training the model, further analysis was conducted during the training process for the action recognition module. The datasets were categorized by action type and split into training and test sets in a 4:1 ratio. The model was subsequently trained to completion, with the primary focus on analyzing the classification results for ten distinct actions. The performance outcomes for these ten classes are presented in **Figure 4**.

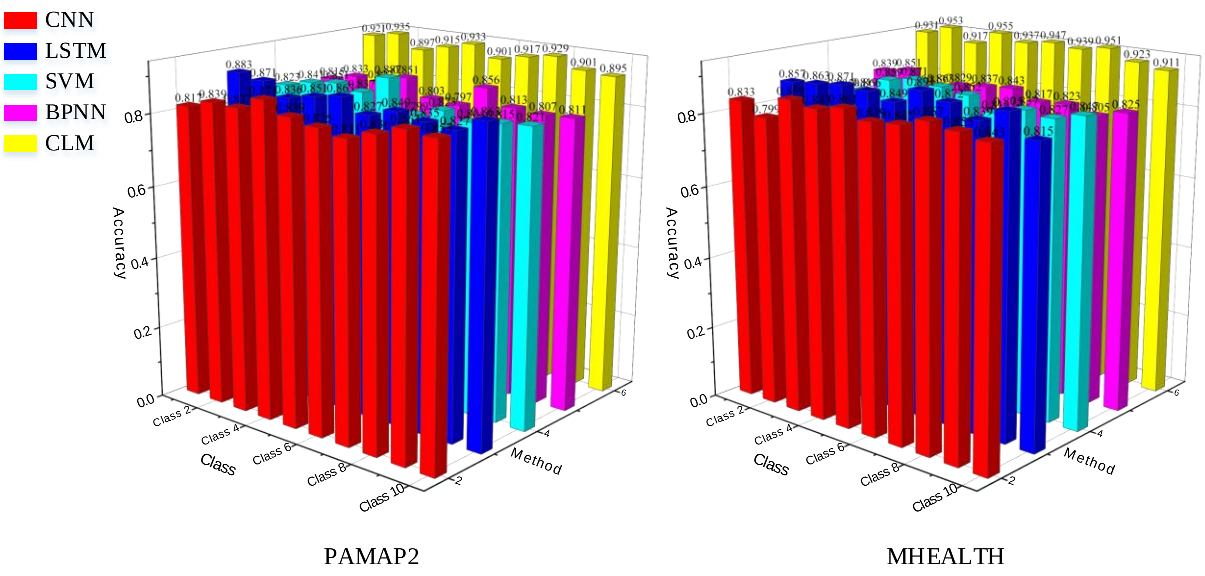

Figure 4. The recognition result among different categories.

The results depicted in **Figure 4** demonstrate that, across the two datasets, the combination of CNN and LSTM methods significantly enhances feature representation, yielding performance metrics that substantially surpass those of traditional machine learning approaches. To provide a more intuitive assessment of the model's comprehensive capabilities in action recognition, a detailed analysis of its performance across various action categories was conducted. The aggregated results of this analysis are illustrated in **Figure 5**.

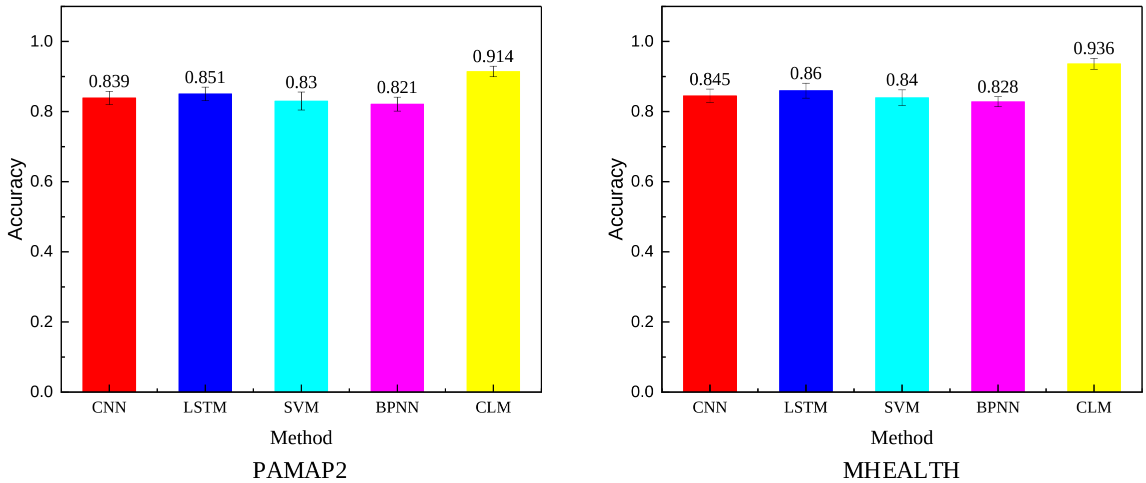

Figure 5. The average recognition results on both datasets.

As illustrated in **Figure 5**, the average recognition accuracy of the CLM framework across the ten categories exceeds 0.9, with minimal error margins, highlighting its robust performance. In contrast, models utilizing only CNN, LSTM, or traditional machine learning methods exhibit recognition accuracies ranging from 0.82 to 0.85 due to the limited representation capacity of single-modal data. These results underscore the significance of combining CNN and LSTM, as this integration effectively enhances the model's feature representation, thereby substantially improving recognition performance.

4.3. Health monitoring result and analysis

Following the performance evaluation of the O&M recognition model, an analysis and comparison of anomaly recognition were conducted. The comparison results across the two datasets are presented in **Figure 6**.

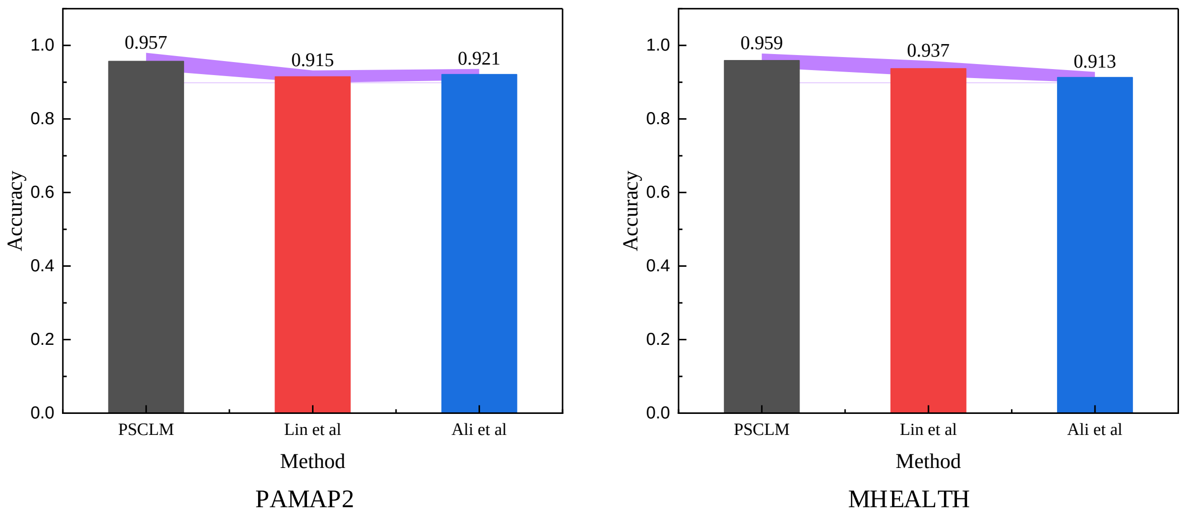

Figure 6. The abnormal states recognition on both datasets.

As shown in **Figure 6**, the proposed method achieves superior performance in identifying anomalous states, with results nearing 0.95 for both datasets. In contrast, Lin's method and Ali's method achieve slightly over 0.9, demonstrating that the approach presented outperforms these existing methods in anomaly detection. To further evaluate the model, we optimized the hyperparameter β and analyzed the performance of different parameters and modules, including the various ablation models described in Section 4.1. The results of this analysis are illustrated in **Figure 7**.

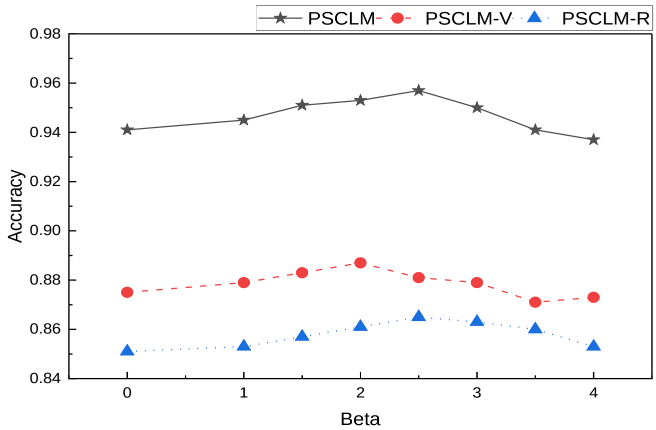

Figure 7. The ablation experiment result with different β.

As illustrated in **Figure 7**, the anomaly recognition accuracy of the three models compared in the ablation experiments initially increases with the rise in the hyperparameter β. However, at later stages, a slight decrease in overall accuracy is observed. Based on this trend, β = 2.5 was identified as the optimal value for further analysis. Additionally, the results reveal that the recognition accuracy of the PSCLM-R model, which excludes motion recognition features, is notably lower, averaging around 0.85. This performance is significantly inferior to the full PSCLM framework proposed, underscoring the importance of motion recognition features.

To further investigate the characteristics of physiological signals, we analyzed heart rate variations under different exercise intensities. The results of this analysis are presented in **Figure 8**.

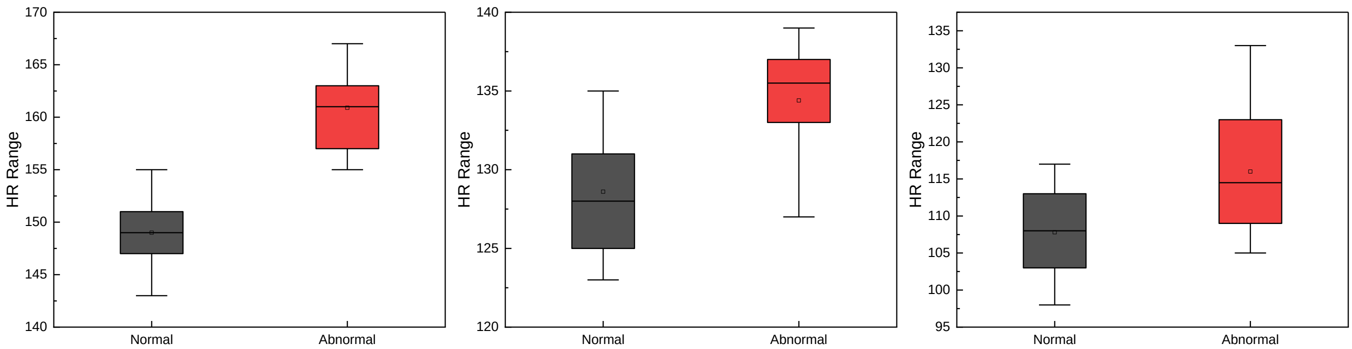

Figure 8. The HR range among different motion intensities.

The heart rate variations depicted in **Figure 8** reveal that while high-intensity exercises such as cycling and running exhibit noticeable heart rate gaps, other exercise types show overlapping heart rate ranges, making it challenging to directly distinguish abnormalities. This highlights the critical role of action recognition in health monitoring, as evidenced by the comparison between the PSCLM and PSCLM-R models.

The inability to rely solely on heart rate for accurate anomaly detection, due to significant individual differences in physical fitness, underscores the necessity of integrating multiple signals and employing multi-level data fusion. Such an approach is key to addressing these challenges and achieving precise identification of movement states and associated health conditions.

5. Discussion

As a core institution for training future law enforcement officers, the Police Academy places significant emphasis on students' physical training, movement skill learning, and health monitoring as key areas of daily teaching and management. By integrating action recognition with physiological signal monitoring, the CLM model and the PSCLM framework can provide scientific support for the training and task execution of law enforcement personnel. The CLM model and the PSCLM health monitoring framework proposed in this study exhibit substantial advantages in action recognition and health status monitoring. For action recognition tasks, the CLM model, which integrates CNN and LSTM architectures, leverages the strengths of CNN for spatial feature extraction and the capacity of LSTMs to model temporal dependencies. The CNN effectively extracts spatial features from inertial sensor data and captures local patterns across various axial signals, ensuring robust performance. The PSCLM health anomaly monitoring framework demonstrates innovation in multimodal data fusion and health state anomaly detection. By applying modal decomposition to ECG signals, the framework extracts frequency-specific ECG features, integrates these with motion recognition outputs and inertial data, and achieves deep feature fusion across multiple levels. This comprehensive approach addresses the limitations of single-modality anomaly detection and effectively captures the interplay between physiological and behavioral data, enabling more precise monitoring of abnormal health state changes. Ablation experiments further validated the critical role of action recognition features in the PSCLM model's performance. Results indicated that incorporating motion features significantly enhances the accuracy of health state anomaly detection. This finding underscores that action information not only reflects behavioral dynamics but also has a strong correlation with health status. Particularly when physiological signals are susceptible to interference, action features provide an additional layer of discriminative power, reinforcing the robustness and reliability of the proposed framework.

By leveraging the strengths of CNN and LSTM, the CLM model achieves high-precision classification of complex motion signals, making it suitable not only for sports health management but also for rehabilitation training, personalized exercise program design, and dynamic behavioral analysis. The PSCLM framework extends this capability by integrating physiological signals (e.g., ECG and HR) with motion recognition features, enabling the deep fusion of multi-level features. This integration significantly enhances the accuracy and robustness of health abnormal state detection. The application of the CLM model and the PSCLM framework provides scientific support for the health management, training optimization, and task execution of police officers, enhancing the efficiency and safety of law enforcement work. In addition, the PSCLM framework is particularly well-suited for applications in chronic disease monitoring, stress assessment, elderly health management, and health alert functions in smart wearable devices, offering valuable insights for the future evolution of health monitoring technologies. In practical implementations, it is crucial to ensure the precision of sensor data collection to minimize errors caused by signal noise or device interference. Furthermore, the model's generalization ability across diverse populations should be carefully considered, with personalized adjustments to enhance its applicability. While the PSCLM framework exhibits high accuracy and robustness, its deployment on mobile and wearable devices poses challenges due to computational and energy constraints. Efficient real-time inference is essential for practical applications, particularly in resource-limited environments. To address this, future work will focus on optimizing the model through lightweight architectures (e.g., MobileNet, EfficientNet) and compression techniques (e.g., pruning, quantization) to reduce computational costs. Furthermore, knowledge distillation strategies will be explored to balance model efficiency and predictive performance. These improvements will enhance PSCLM's adaptability for edge computing scenarios, ensuring its effectiveness in real-world applications.

In summary, the proposed framework provides an innovative technological pathway for motion recognition and health monitoring, with promising future applications in telemedicine, health management, and smart wearable devices.

6. Conclusion

In this paper, we propose a health monitoring framework—PSCLM—that integrates CNN, LSTM, and modal decomposition techniques to address the challenges of motion recognition and health abnormality monitoring. This framework aims to overcome the limitations of traditional methods in processing multimodal data and monitoring health states. Initially, a CLM model is developed by preprocessing and extracting features from inertial sensor data using CNN and LSTM, achieving accurate recognition of multi-class actions. Subsequently, ECG signals undergo modal decomposition to extract multi-level time-frequency features, which are fused with motion recognition and heart rate features within a multimodal framework for health anomaly monitoring. Experimental results demonstrate the superiority of the PSCLM framework over traditional methods and single deep learning models. On the PAMAP2 and MHEALTH datasets, the framework achieves an action classification accuracy exceeding 0.9 and an abnormal state detection accuracy of 0.95, showcasing its robustness and adaptability.

This study focused on heart rate and electrocardiogram (ECG) signals collected from wearable devices. While signals such as EEG and blood oxygen levels could enhance physiological monitoring, their accuracy in wearable settings remains a challenge due to sensor limitations and motion artifacts. Therefore, we prioritized more reliable signals in this study. In future research, we will explore the integration of additional physiological signals to improve health monitoring, especially in high-stress or extreme environments. In summary, future research will not only drive the in-depth development of multimodal data fusion technology but also provide robust technical support for health monitoring, task optimization, and dynamic management within law enforcement operations, thereby enhancing the intelligence level and overall safety of policing work. Meanwhile, the extensive application potential of this framework in various health monitoring scenarios will be thoroughly explored, including telemedicine, rehabilitation training, and elderly health management. These research efforts and applications will collectively promote the development of an intelligent and flexible health monitoring system, laying a solid theoretical foundation and practical experience for future innovations in sports health management and health anomaly monitoring.

Author contributions: Conceptualization, GX and WL; methodology, GX; software, GX; validation, GX, WL and XL; formal analysis, GX; investigation, GX; resources, GX; data curation, WL; writing—original draft preparation, GX; writing—review and editing, GX; visualization, XL; supervision, XL; project administration, XL; funding acquisition, XL. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Zhejiang Provincial universities major humanities and social science research planning key project “15-minute Public Service Circle” to help common prosperity typical case and inspiration Research (Grant No. 2024GH020).

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. He Z, Li W, Salehi H, et al. Integrated structural health monitoring in bridge engineering. Automation in Construction. 2022; 136: 104168. doi: 10.1016/j.autcon.2022.104168
2. Dong CZ, Catbas FN. A review of computer vision–based structural health monitoring at local and global levels. Structural Health Monitoring. 2020; 20(2): 692-743. doi: 10.1177/1475921720935585
3. Sujith AVLN, Sajja GS, Mahalakshmi V, et al. Systematic review of smart health monitoring using deep learning and Artificial intelligence. Neuroscience Informatics. 2022; 2(3): 100028. doi: 10.1016/j.neuri.2021.100028
4. Jayawickrema UMN, Herath HMCM, Hettiarachchi NK, et al. Fibre-optic sensor and deep learning-based structural health monitoring systems for civil structures: A review. Measurement. 2022; 199: 111543. doi: 10.1016/j.measurement.2022.111543
5. Jefiza A, Pramunanto E, Boedinoegroho H, et al. Fall detection based on accelerometer and gyroscope using back propagation. 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). 2017: 1-6. doi: 10.1109/eecsi.2017.8239149
6. Ishii S, Nkurikiyeyezu K, Luimula M, et al. ExerSense: Real-Time Physical Exercise Segmentation, Classification, and Counting Algorithm Using an IMU Sensor. Activity and Behavior Computing. 2020; 204: 239–255.
7. Dawar N, Kehtarnavaz N. Action Detection and Recognition in Continuous Action Streams by Deep Learning-Based Sensing Fusion. IEEE Sensors Journal. 2018; 18(23): 9660-9668. doi: 10.1109/jsen.2018.2872862
8. Kui L, Chen C, Jafari R, et al. Fusion of Inertial and Depth Sensor Data for Robust Hand Gesture Recognition. IEEE Sensors Journal. 2014; 14(6): 1898-1903. doi: 10.1109/jsen.2014.2306094
9. Li H, Shrestha A, Fioranelli F, et al. Multisensor data fusion for human activities classification and fall detection. 2017 IEEE SENSORS. Published online October 2017. doi: 10.1109/icsens.2017.8234179

10. Ehatisham-Ul-Haq M, Javed A, Azam MA, et al. Robust Human Activity Recognition Using Multimodal Feature-Level Fusion. IEEE Access. 2019; 7: 60736-60751. doi: 10.1109/access.2019.2913393
11. Radu V, Tong C, Bhattacharya S, et al. Multimodal Deep Learning for Activity and Context Recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018; 1(4): 1-27. doi: 10.1145/3161174
12. Sadeghi R, Banerjee T, Hughes JC, et al. Sleep quality prediction in caregivers using physiological signals. Computers in Biology and Medicine. 2019; 110: 276-288. doi: 10.1016/j.compbiomed.2019.05.010
13. Nho YH, Lim JG, Kwon DS. Cluster-Analysis-Based User-Adaptive Fall Detection Using Fusion of Heart Rate Sensor and Accelerometer in a Wearable Device. IEEE Access. 2020; 8: 40389-40401. doi: 10.1109/access.2020.2969453
14. Kim Y, Jeung J, Song Y, et al. A Wearable System for Heart Rate Recovery Evaluation with Real-Time Classification on Exercise Condition. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Published online November 1, 2021: 7609-7612. doi: 10.1109/embc46164.2021.9629673
15. Xiao N, Yu W, Han X. Wearable heart rate monitoring intelligent sports bracelet based on Internet of things. Measurement. 2020; 164: 108102. doi: 10.1016/j.measurement.2020.108102
16. Spathis D, Perez-Pozuelo I, Brage S, et al. Learning Generalizable Physiological Representations from Large-scale Wearable Data. arXiv preprint arXiv:2011.04601. 2020; 1-6.
17. Chong W, Kim S, Yu C, et al. Analysis of Health Management Using Physiological Data Based on Continuous Exercise. International Journal of Precision Engineering and Manufacturing. 2021; 22(5): 899-907. doi: 10.1007/s12541-021-00503-3
18. Gao ZK, Li YL, Yang YX, et al. A recurrence network-based convolutional neural network for fatigue driving detection from EEG. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2019; 29(11). doi: 10.1063/1.5120538
19. Porumb M, Stranges S, Pescapè A, et al. Precision Medicine and Artificial Intelligence: A Pilot Study on Deep Learning for Hypoglycemic Events Detection based on ECG. Scientific Reports. 2020; 10(1). doi: 10.1038/s41598-019-56927-5
20. Shi X, Wang Z, Zhao H, et al. Threshold-Free Phase Segmentation and Zero Velocity Detection for Gait Analysis Using Foot-Mounted Inertial Sensors. IEEE Transactions on Human-Machine Systems. 2023; 53(1): 176-186. doi: 10.1109/thms.2022.3228515
21. Lin F, Wang Z, Zhao H, et al. Adaptive Multi-Modal Fusion Framework for Activity Monitoring of People With Mobility Disability. IEEE Journal of Biomedical and Health Informatics. 2022; 26(8): 4314-4324. doi: 10.1109/jbhi.2022.3168004
22. Yang S, Yang H, Li N, et al. Short-Term Prediction of 80–88 km Wind Speed in Near Space Based on VMD–PSO–LSTM. Atmosphere. 2023; 14(2): 315. doi: 10.3390/atmos14020315
23. Bollampally A, Kavitha J, Sumanya P, et al. Optimizing Edge Computing for Activity Recognition: A Bidirectional LSTM Approach on the PAMAP2 Dataset. Engineering, Technology & Applied Science Research. 2024; 14(6): 18086-18093. doi: 10.48084/etasr.8861
24. Triantafyllidis A, Kondylakis H, Katehakis D, et al. Deep Learning in mHealth for Cardiovascular Disease, Diabetes, and Cancer: Systematic Review. JMIR mHealth and uHealth. 2022; 10(4): e32344. doi: 10.2196/32344
25. Ali F, El-Sappagh S, Islam SMR, et al. An intelligent healthcare monitoring framework using wearable sensors and social networking data. Future Generation Computer Systems. 2021; 114: 23-43. doi: 10.1016/j.future.2020.07.047