International Journal of Innovative Computing, Information and Control
Volume **21**, Number **4**, August **2025**

ICIC International ©2025 ISSN 1349-4198
pp. **1149–1166**

DEEP BI-GRU WITH MULTI-HEAD ATTENTION MECHANISM FOR HUMAN ACTIVITY RECOGNITION WITH WEARABLE SENSORS IN INDUSTRIAL ENVIRONMENT

Haiquan Wang[1,*], Jiabo Zhai[2], Yueyi Yang[1], Xiaobin Xu[3,4], Shengjun Wen[1]
Jinxia Wen[2] and Yabo Hu[2]

[1]Zhongyuan-Petersburg Aviation College
[2]School of Automation and Electrical Engineering
Zhongyuan University of Technology
No. 41, Zhongyuan Middle Road, Zhengzhou 450007, P. R. China
{ 2022106230; yangyy; wsj; 2023006218; 2024106042 }@zut.edu.cn
[*]Corresponding author: wanghq@zut.edu.cn

[3]Department of Automation
[4]The Belt and Road Information Research Institute
Hangzhou Dianzi University
No. 1158, 2nd Avenue, Baiyang Sub-street, Qiantang District, Hangzhou 310018, P. R. China
xuxiaobin1980@hdu.edu.cn

Received November 2024; revised March 2025

Abstract. *Misoperations or unsafe activities of workers in industrial environment are one of the main potentials of accidents. To recognize the abnormal behavior in time, a novel human activity recognition (HAR) model combined with multiple attention modules based on a bidirectional gated recurrent unit (Bi-GRU), is proposed in this paper. Here multi-head attention (MHA) mechanism is used to extract more comprehensive or representative features from inertial measurement unit (IMU) signals, and the residual connection is designed to avoid gradient disappearance. Based on the features extracted above, a deep Bi-GRU with a self-attention (SA) module is introduced to further optimize the features which makes the model possess a more comprehensive understanding of the changes in actions and improve the accuracy of HAR. To verify the effectiveness of the proposed model, a workers' climbing behavior dataset (WCBD) in the production environment was constructed, and corresponding experiments were executed. The results on the public daily activity datasets PAMAP2 and WCBD show that the recognition accuracies are 98.06% and 95.86% respectively. Compared with other advanced models, the model can improve recognition accuracy, with the accuracy reaching a maximum increase of 8.03% and 3.41%, respectively.*

Keywords: Human activity recognition, Deep learning, Multi-head attention, Wearable sensors, Workplace safety

1. **Introduction.** In nowadays, the manufacturing industry still requires a significant number of personnel to perform necessary operations, and over 85% of accidents are caused by misoperations or unsafe human behaviors [1]. With the development of the Internet of Things (loT) technology and Industry 4.0, for preventing potential safety accidents, ensuring workers' safety, and maintaining a safe and productive industrial environment, real-time abnormal behavior recognition is essential. As manual safety monitoring is time-consuming and ineffective [2], there is a growing need for vision-based methods or wearable sensor-based methods, which could achieve human activity recognition (HAR) in industrial environments. Vision-based methods need the workers to be always visible to the

DOI: 10.24507/ijicic.21.04.1149

camera and are sensitive to background [3], and low power consuming wearable sensors methods are more suitable for industrial environments, because they are unaffected by lighting variations and visual obstructions [3,4]. Wearable sensors, including but not limited to motion, biological, and pressure sensors, have been extensively studied for their diverse applications in HAR. Han et al. [5] utilized motion sensors and biological sensors to collect acceleration and heart rate data, achieving 98.19% accuracy in classifying two types of daily activities through data fusion. In another study, Zhang et al. [6] developed a gesture recognition system based on pressure sensors, which attained 94.6% recognition accuracy for seven common hand gestures. Among them, the inertial measurement unit (IMU) is most widely used, and the acceleration or gyroscope signals are collected from different body segments such as the wrists, shoulders, knees, or other limbs and data-driven models are used for HAR. IMU-based methods have already many successful application cases in industrial environments. Yan et al. [7] developed an IMU-based motion warning system that helped construction workers to enhance their self-management capabilities by monitoring and warning of hazardous postures in high-altitude operations in real time. A sensor-based motion recognition system was constructed by Zhu and Hwang [8] to assess workers' unsafe behaviors in real time, which provided an effective digital tool for safety management in the construction industry.

For HAR, traditional machine learning strategies, such as support vector machines and decision trees have already been widely applied [9,10], but the extracted features depend heavily on expert experience and corresponding parameters need to be adjusted by trial and error, and the optimal classification performance cannot be guaranteed. With the development of deep learning technology, different kinds of deep network strategies which could learn high-level features automatically and represent the data in a more reasonable way, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), have been designed and applied for HAR [11]. Boufama et al. [12] designed a model based on CNNs and RNNs for the task recognition of operatives. Despite the excellent performance of CNN in HAR, its high computational complexity makes it difficult to meet the requirements of real-time applications in human action recognition. In contrast, network structures such as long short-term memory (LSTM) and gated recurrent unit (GRU) can improve the efficiency and accuracy of real-time human action recognition simultaneously. Mekruksavanich et al. [13] collected data from IMUs at multiple body locations of workmen and applied LSTM to recognizing their daily activities, including 16 actions such as roll painting, climbing stairs, and crouching floor. Li et al. [14] adopted LSTM to recognize the daily activities of workmen, which include walking, running, climbing, sitting, lying down, and standing. LSTM and GRU have advantages in processing time-series data, and GRU has a lower computational complexity compared with LSTM [15], which is very suitable for identifying worker activities in industrial environments. However, GRU can only process the information of past and current moments and is unable to handle the information of the next moment. In contrast, the bidirectional gated recurrent unit (Bi-GRU) has significant advantages in processing complex time-series inputs in both directions, which makes the model more robust and reliable in various applications. However, Bi-GRU faces problems such as gradient vanishing or explosion when dealing with long sequences, and it is difficult to capture long-term dependencies.

To address these issues, the attention mechanisms have been widely applied in deep learning. The attention mechanism can focus on key features and suppress useless features without incurring additional computational costs [16,17]. Mekruksavanich et al. [18] developed a CNN-RNN model with the convolutional block attention module (CBAM) and successfully recognized 7 types of actions of sanitation workers, including walking, running,

sweeping the floor, and so on. Zhang and Xu [19] incorporated the spatio-temporal attention mechanism into the CNN-BiGRU network for HAR. The enhanced model achieved an accurate improvement ranging from 3.55% to 8.59% across multiple public datasets. Wang et al. [20] proposed a new model based on GNN with multi-head attention (MHA), which extracts the spatial and temporal features from the graph-structured data. Although the MHA mechanism demonstrates remarkable performance in feature extraction, it encounters challenges in terms of over-parameterization. This over-parameterization not only leads to a larger model size but also increases the risk of overfitting, especially when dealing with limited-scale datasets in real-time HAR environments. Considering the strict requirements for model generalization and the limited data availability in real-time HAR, it becomes essential to design a new model to balance the performance and generalization ability.

Based on previous analyses, a new HAR model using deep Bi-GRU and multiple attention mechanism is proposed, where multiple attention mechanism is composed of self-attention (SA) and MHA. Each MHA attention head acts as an independent SA, processing data from different views and combining results for multi-dimensional learning. This unique architecture of MHA, along with its integration with SA, enables the model to capture diverse features, extracting key information from time-series data more effectively.

The rest of the paper is organized as follows. Section 2 describes the model proposed in this paper. Section 3 presents experimental settings and results. Section 4 demonstrates the practical application of the model. The conclusion is provided in Section 5.

2. **Proposed Model.**

To improve the performance of HAR, the MHA mechanism is used to extract more comprehensive feature information from multiple sensor signals, and SA-BiGRU is developed to extract temporal features effectively. The proposed model, including MHA and SA mechanisms, not only captures richer feature information at the front end of the model but also reduces the negative influence of redundant features on recognition results.

2.1. **Self-attention mechanism.**

The SA mechanism is a technique for capturing interrelationships among elements in sequential data. Its core principle is to dynamically assign the importance of different elements by calculating the attention weights between each element in the sequence and other elements, thus generating a context-aware representation for each element. The calculation process of SA is shown in Equations (1)-(3).

First of all, the input feature maps are transformed into three distinct matrices, namely Q (a query matrix), K (a key matrix), and V (a value matrix). The similarity scores between Q and K vectors are calculated and which elements of the input will receive more attention could be determined.

$$e_i = F(Q, K) \tag{1}$$

Here Q is used to extract the information of interest from the input data, K determines the importance of each position, e_i indicates the ith attention scores, and $F(\cdot)$ indicates the function that calculates the similarities between Q and K. Then the scores are transformed into a probability distribution through the Softmax function expressed as Equation (2), and the transformed scores emphasize the weights of significant information, enabling the model to focus on important information.

$$m_i = soft\max(e_i) = \frac{\exp(e_i)}{\sum_{j=1}^{N} \exp(e_j)} \tag{2}$$

Thirdly, a weighted sum of V vectors is performed based on the transformed scores as shown in Equation (3). By multiplying these vectors with their corresponding weights and

summing them up, the attention mechanism yields the final output, effectively capturing the essence of the input data by highlighting its salient features.

$$Attention((K,V),Q) = \sum_{i=1}^{N} m_i v_i \tag{3}$$

where $i, j \in [1, N]$ are the positions of the vector sequence, V indicates the sequence of value vector, v_i indicates the ith value vector, and m_i indicates the ith weight coefficient, which is dynamically generated by the SA mechanism.

The process of the SA mechanism is shown in Figure 1; obviously, the attention mechanism helps the model decide which parts of the input are most important and then combines those parts in a meaningful way to improve its understanding and processing.

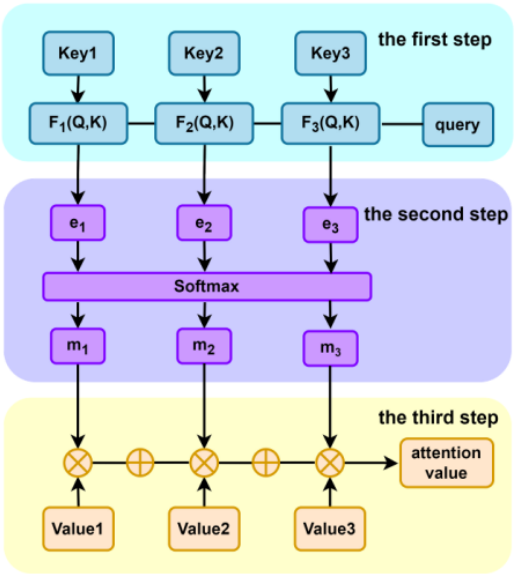


FIGURE 1. Architecture of SA mechanism

2.2. **Multi-head attention mechanism.** To improve the model's ability to capture long-range dependencies, and understand input information comprehensively, the MHA mechanism is introduced.

First of all, each head in the MHA has its own learnable parameters as shown in Equation (4), which enables each head to perform the attention operation. Therefore, each head can be regarded as an SA mechanism.

$$Q_i = QW_q^i, \quad K_i = KW_k^i, \quad V_i = VW_v^i \tag{4}$$

where W_q^i, W_k^i, W_v^i denote the learnable parameters. Q_i denotes the query matrix of the ith attention head, K_i denotes the key matrix of the ith attention head, and V_i denotes the value matrix of the ith attention head.

Then, the output of each attention head is calculated, as shown in Equations (5) and (6). To avoid problems such as gradient vanishing or exploding during the training phase, scaled dot-product attention is adopted by MHA to calculate the similarity between Q and K. These scores are scaled down by dividing by the square root of the dimensionality of the K vector. After scaling, the attention scores are normalized through the softmax function. Finally, the normalized scores are multiplied by the corresponding V.

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

$$head_i = Attention(Q_i, K_i, V_i) \tag{6}$$

where d_k denotes the dimensionality of the K vectors, i denotes the number of heads in MHA, and $head_i$ denotes the features learned from the ith feature subspace.

Thirdly, as shown in Equation (7), the output of each head is combined and linear transformed to obtain the final output. It is a vector matrix that contains multi-dimensional information.

$$Multi = Concat\left(head_1, \ldots, head_n\right) W^0 \tag{7}$$

where W^0 denotes weight. This process not only increases the model's capacity to capture complex relationships but also enhances its ability to generalize across different types of data input. The process of MHA mechanism is shown in Figure 2. Through multiple parallel SA heads, MHA can capture the complex relationships between different positions in the input sequence. This approach not only improves the model's representational ability but also enhances its ability to capture long-range dependencies.

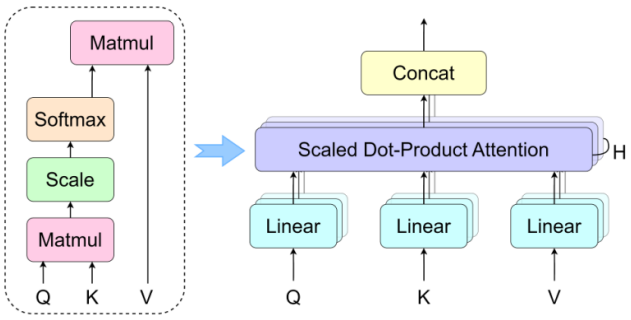


FIGURE 2. MHA mechanism

MHA can extract comprehensive features from multiple perspectives, and it is more suitable for extracting high value features from multiple sensors signals at the input end of the model. After the feature extraction of MHA, SA can further optimize these features and highlight key information. Therefore, the proposed model can capture rich feature information in the early stages and more effectively focus on important features in the subsequent stages. The proposed model not only can obtain richer input information by MHA but also capture better features by SA, which can improve the recognition performance of the model.

2.3. **SA-BiGRU.** A gating mechanism is employed in GRU, which can obtain the changing state of the input data and can be denoted as

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{8}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{9}$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \Theta h_{t-1}, x_t] + b_h) \tag{10}$$

$$h_t = (1 - z_t)\Theta h_{t-1} + z_t \Theta \tilde{h}_t \tag{11}$$

where x_t is the input data at time t, σ is the sigmoid activation function, h_{t-1} is the hidden state passed down from the previous node, h_t is the hidden state passed to the next node, $\tilde{h}_t$ is the candidate hidden state, W is a weight matrix, b is a bias vector, r_t is the reset gate, which discards irrelevant information by controlling the input of the hidden state in the previous time, and z_t is the update gate, which is used for the forgetting and selecting memory, which controls the process of updating information to the new state h_t or the candidate state $\tilde{h}_t$. As shown in Figure 3, feature representations can be extracted from both forward and backward time-series data by Bi-GRU, which can be denoted by

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right) \tag{12}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t-1}\right) \tag{13}$$

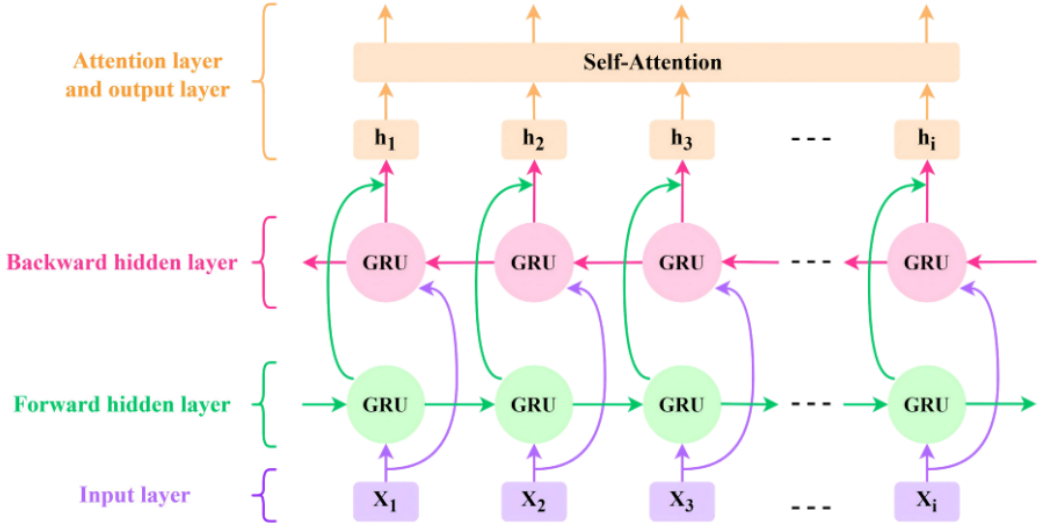


FIGURE 3. Structure of Bi-GRU with self-attention mechanism

The forward GRU is used to capture the information from historical data while the backward GRU is employed to obtain the information from the activity data of the next node. To enhance the performance of Bi-GRU, the SA is introduced to focus on key information and reduce attention weights of redundant information, which is denoted by

$$h_t = s_t\vec{h}_t + d_t\overleftarrow{h}_t + b_t \tag{14}$$

$$y = h_t \cdot soft\max(G_a \cdot \tanh(h_t)) \tag{15}$$

where s_t is the weight matrix of the forward GRU output; d_t is the weight matrix of the backward GRU output; b_t is the bias of the weight matrix; G_a is the learnable weight matrix during the training in the attention layer. The introduction of the SA mechanism not only enhances the Bi-GRU network's ability to extract features but also improves its adaptability and accuracy in complex environments.

The proposed model is shown in Figure 4, and the MHA mechanism extracts comprehensive feature information from the raw data from multiple perspectives. Subsequently, a residual connection is designed to combine the raw data with the features optimized by MHA, ensuring that the Bi-GRU can effectively extract more diverse features. After that, the data combined through the residual connection is input into the Bi-GRU. With the help of the SA module, the Bi-GRU captures the long-term dependencies in the time-series data to further optimize the features. The SA module is used to emphasize key information and filter out irrelevant information to assist in extracting high-quality information. Finally, the fully connected layer and the Softmax function are used to predict the identification results. In the complex and dynamic environment of industrial production, coherence and changing trends of workers' actions can be more accurately understood by the proposed model, which can effectively recognize the unsafety activities and reduce misjudgments and thereby enhance the safety and efficiency of industrial production.

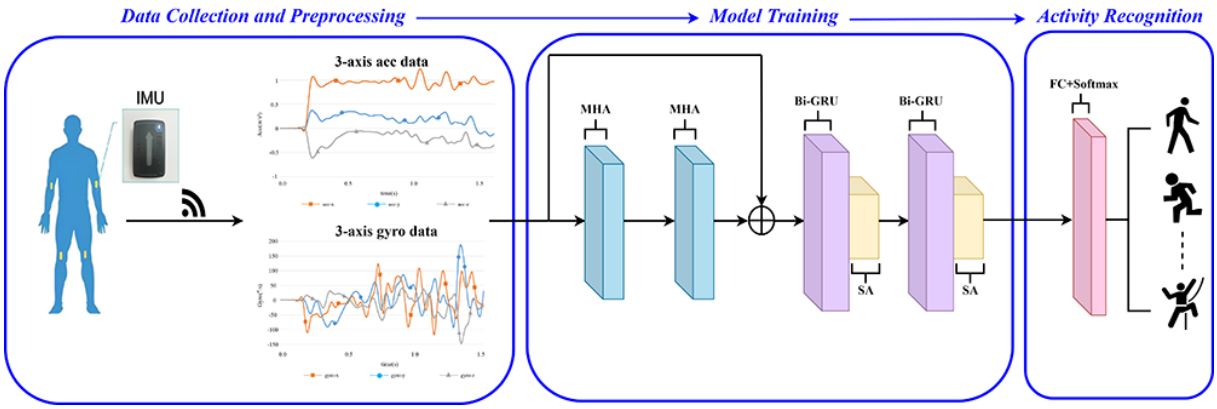


FIGURE 4. The framework of the proposed strategy

3. **Experimental Results and Analysis.** To validate the performance of the proposed model, experiments are carried out on PAMAP2 datasets [21]. A workers' climbing behavior dataset (WCBD) is constructed to verify the generalization ability of the model and address the wrong activities of workers in industrial environments.

3.1. **Datasets.** The details of the two datasets are shown in Table 1. The PAMAP2 dataset includes 12 daily activity data recorded from 9 participants. The subjects (one female and eight males) are between 23 and 30 years of age and had body mass indexes between 22.49 and 27.73 kg/m^2. Three IMUs are worn on each subject's wrist of the dominant arm, the chest, and the dominant side's ankle, respectively. The dataset contains 12 different activities, including lying, sitting, standing, walking, and so on.

TABLE 1. Details of the datasets

Dataset	Sensor position	Data type	Types of actions
PAMAP2	Arm Chest Ankle	3-axis-acc 3-axis-gyro	Lying (L), Sitting (SI), Ironing (I), Running (R) Cycling (CY), Standing (ST), Walking (WL) Rope Jumping (J), Nordic Walking (NW), Ascending Stairs (US), Descending Stairs (DS), Vacuum Cleaning (CL)
WCBD	Limbs	3-axis-acc 3-axis-gyro	Climbing (wrong) Climbing (correct)

The WCBD dataset is constructed within a real-world industrial environment, and fourteen subjects are selected who wear four IMUs on their limbs. The dataset categorizes climbing activities into two groups: wrong climbing action and correct climbing action. If only one limb leaves the ladder at any time during the climbing process, it is considered a correct climbing action and if there is a moment during the climbing process when multiple limbs leave the ladder simultaneously, it is judged as a wrong climbing action. Taking the x-axis gyroscope data of limbs as an example, the data acquisition process is illustrated in Figure 5, and the key information of these two actions is marked with black dashed boxes. When performing a correct climbing action, only the gyroscope signals of one limb fluctuate greatly each time, and the other signals are relatively stable. When performing a wrong climbing action, the gyroscope signals of two limbs fluctuate greatly each time, and the other signals are relatively stable.

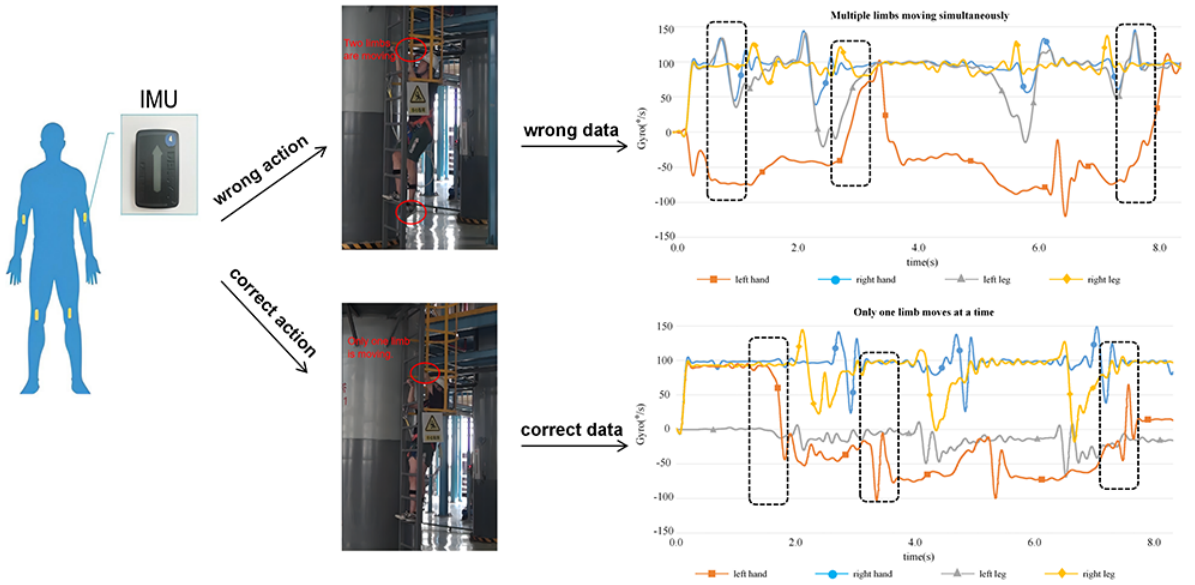
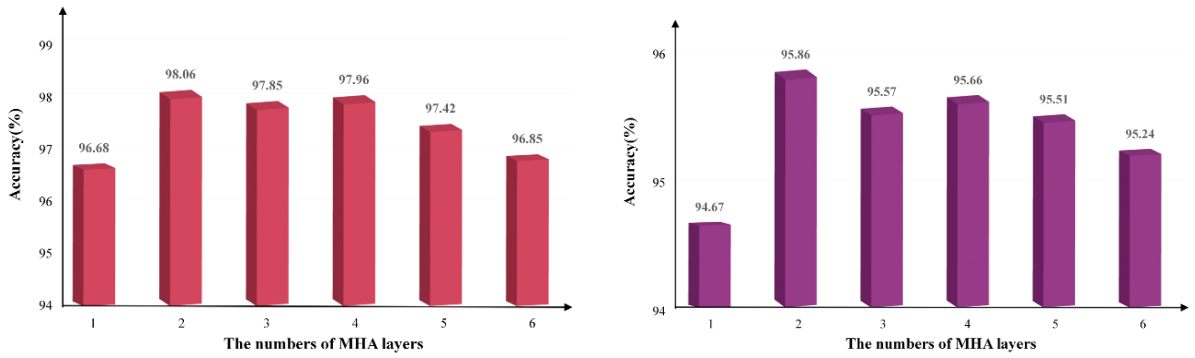


FIGURE 5. Data acquisition process of WCBD dataset

3.2. **Experimental setting.** The numbers of Bi-GRU layers, SA layers, and MHA layers have impacts on the performance of the proposed model. To obtain the optimal values of the above parameters, experiments were conducted on the WCBD dataset and the PAMAP2 dataset. For the MHA layers, the number of MHA layers is set as 1, 2, 3, 4, 5, and 6, respectively. The experimental results are shown in Figure 6, the highest accuracy is achieved when the number of MHA is 2, and the accuracy of the model decreases as the number of MHA increases. The reason is that more MHA layers will lead to a more complex model, making model training difficult and requiring more training data.



(a) (b)

FIGURE 6. Influence of the number of MHA layers: (a) PAMAP2; (b) WCBD

The numbers of Bi-GRU layers are set to be 1, 2, and 3, respectively, each layer comprises 128 Bi-GRU units, and SA is integrated into each Bi-GRU layer. As shown in Figure 7, the results of the experiments on the numbers of Bi-GRU layers and SA layers are presented. The experimental results indicate that the proposed model achieves the highest accuracy when the numbers of Bi-GRU layers and SA layers are both 2. Specifically, the highest accuracy on PAMAP2 dataset is 98.06%, and 95.86% on WCBD dataset. The length of the sliding window (WL) has a great impact on model performance. A too-short window may fail to capture the complete data pattern and lead to a decrease in accuracy,

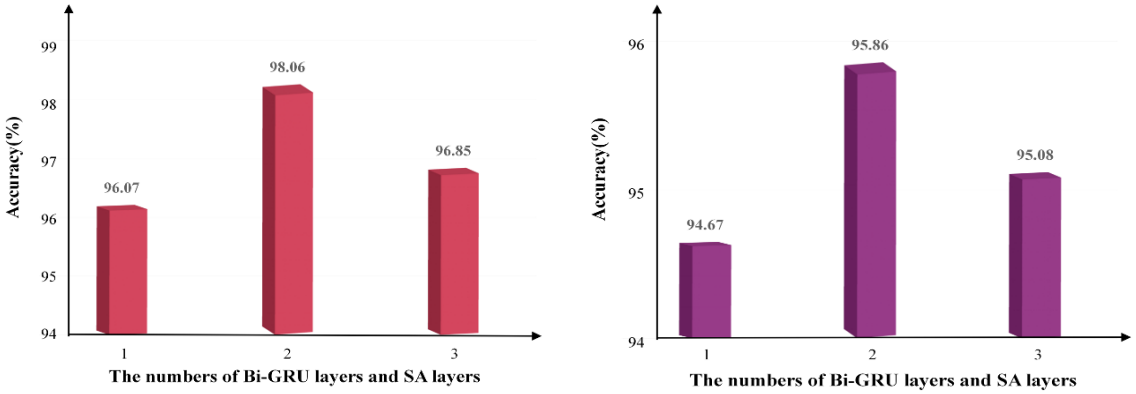


FIGURE 7. Influence of the number of Bi-GRU layers and SA layers: (a) PAMAP2; (b) WCBD

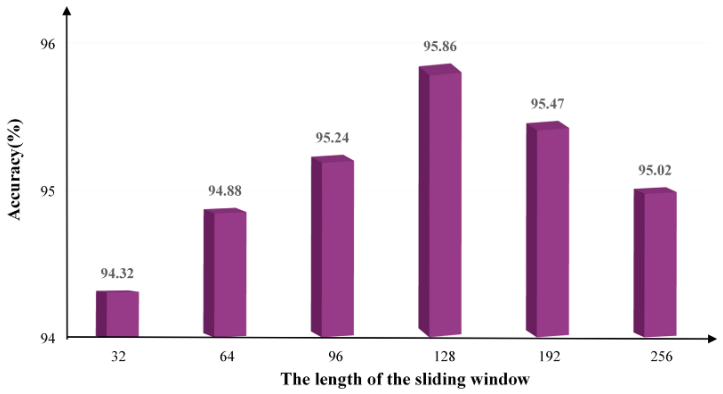


FIGURE 8. Influence of the length of sliding window

while a too-long window will introduce redundant information and noise, increase computational cost, and reduce the number of effective samples. For the PAMAP2 dataset, the setting of WL is the same as that in [24,25], and the WL is 171. For the WCBD dataset, the values were set as 32, 64, 96, 128, 192, and 256, respectively, and a series of experiments were conducted to select the best effect. As shown in Figure 8, it is indicated that model accuracy reaches peak value when the WL is set to be 128.

The parameter settings of the two datasets are detailed in Table 2, including the sensor sampling rate (SR), WL and so on. The ratio of the training set to the test set for both datasets is 7 : 3. After the overall loss function is calculated, the Adam optimizer is employed for training with an initial learning rate of $1e^{-4}$ and weight decay of $1e^{-5}$. The proposed model is implemented on the TensorFlow 2.1 framework using two NVIDIA 3090 GPUs. The evaluation metrics in the HAR study can be used to adjust the model parameters, thereby enhancing the classification efficacy of the model. This study employed accuracy (ACC), F1-score, and the area under the curve (AUC) as evaluation metrics. ACC can intuitively reflect the proportion of samples correctly predicted by the model in the total number of samples. The F1-score can more comprehensively evaluate the performance of the model in classification. In the case of data imbalance, AUC is a very reliable metric for evaluating the generalization ability of the model.

TABLE 2. Pre-processing of datasets

Dataset	WL	SR	Channels	Samples
PAMAP2	171	100 Hz	40	16490
WCBD	128	128 Hz	24	6853

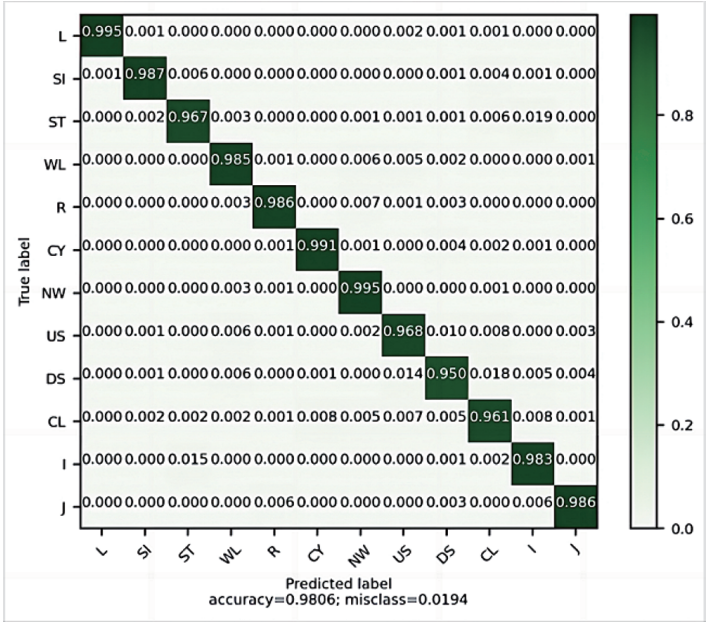


FIGURE 9. Confusion matrix of the proposed model on the PAMAP2 dataset

TABLE 3. Comparison results on the PAMAP2 dataset

Model	ACC (%)	F1 (%)
Wang et al. [20]	96.74	96.33
Wang et al. [22]	93.65	93.78
Tang et al. [23]	94.29	93.75
Yao et al. [24]	95.53	95.51
Lu and Deng [25]	97.11	97.12
Proposed model	**98.06**	**98.13**

3.3. **Experiment analysis.** For the PAMAP2 dataset, the confusion matrix results of the proposed model are shown in Figure 9, and the comparison results with other models are presented in Table 3. From this, it is clear that compared with the advanced models of other authors, the proposed model has the highest accuracy, which demonstrates the superiority of the model.

Wang et al. [20] proposed the MhaGNN model, which mainly combines MHA and GNN. In terms of accuracy, it is 1.32% less than the proposed model. This is mainly because action signals have strong temporal features. The proposed model can extract bidirectional temporal features simultaneously and further optimize key feature information with the help of the attention module. However, MhaGNN may not be able to accurately capture such information. In addition, experiments were conducted to determine the optimal parameters, such as the number of MHA layers, the number of Bi-GRU layers, and the WL. These parameters enable the model to better adapt to the action recognition tasks

in industrial environments. In contrast, MhaGNN has not been fully optimized for industrial environments. The dynamic Gaussian convolution (DgConv) model based on dynamic Gaussian kernel proposed by Wang et al. [22] performs well in certain specific data distributions or relatively simple activity scenarios. However, in complex industrial environments, these advantages are overshadowed by their limitations. DgConv method lacks multi-perspective feature learning ability like MHA and has difficulty in focusing on and adaptively allocating weights to the critical features of different time points and action parts. Therefore, its accuracy is not as good as the proposed model. To improve the performance of CNNs without increasing memory and computational burden, Tang et al. [23] proposed the idea of hierarchical split (HS), although it can enhance multi-scale feature representation, it is less efficient in handling actions with temporal sequences and complexity, and thus its accuracy is 3.77% inferior to that of the proposed model. Compared with the model proposed in this paper, the model proposed by Yao et al. [24] has an accuracy 2.53% lower. This is because their model may not be as comprehensive as Bi-GRU when dealing with complex time-series, and it lacks the ability of the attention mechanism to focus on key features. Compared to the denoising-focused approach of Lu and Deng [25], the proposed model achieves improved performance in complex action recognition by prioritizing key feature extraction. As a result, its accuracy is 0.95% lower than that of the proposed model. However, the proposed denoising method can be easily integrated into existing HAR architecture, providing new ideas for research.

For the WCBD dataset, the results of the confusion matrix are shown in Figure 10. It can be seen that the climbing action recognition accuracy reaches 97.7% and 93.6%, respectively. Obviously, the model can accurately capture the key features during climbing, and it has a high classification level and good generalization ability.

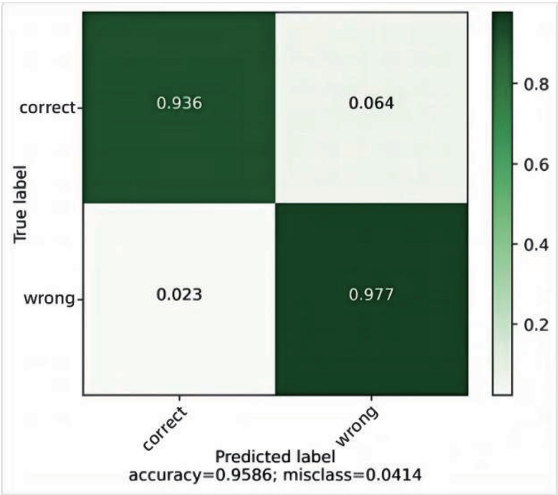


FIGURE 10. Confusion matrix of the proposed model on the WCBD dataset

3.4. **Ablation study.** To validate the effectiveness of each module within the framework and conduct an ablation study, the proposed strategy is compared with the baselines listed below. Through these comparisons, the analysis focuses on how the removal or modification of each module impacts the overall performance of the model, thereby clearly understanding the contribution of each individual component in the framework.

1) GRU: It utilizes a two-layer GRU with each layer comprising 128 GRU units.

2) Bi-GRU: It adopts a two-layer Bi-GRU in which each layer is made up of 128 Bi-GRU units.

3) SA-BiGRU: It makes use of a two-layer Bi-GRU where every layer contains 128 Bi-GRU units, and an SA layer is attached after each Bi-GRU layer.

4) MHA-GRU: It employs a two-layer MHA to extract features and then transmits them to a two-layer GRU architecture through residual connections, with each GRU layer consisting of 128 GRU units.

5) MHA-BiGRU: It utilizes a two-layer MHA to extract features, and subsequently, these features are transmitted to a two-layer Bi-GRU architecture through residual connections. Each of the Bi-GRU layers is composed of 128 Bi-GRU units.

For the PAMAP2 dataset, the comparison results with the baseline models are shown in Table 4, and the results prove the effectiveness of each module of the proposed model. Bi-GRU is better at capturing information in both directions, and it can get more detailed features and make the prediction more accurate. Therefore, the Bi-GRU model has an accuracy of 0.81% higher than that of the GRU. SA can handle complex relationships more effectively, which obtains the important parts and cuts down the disturbing noise, So the accuracy of SA-BiGRU is 0.29% higher than Bi-GRU. The SA-BiGRU model has high recognition accuracy for actions such as "walking" and "running", but its recognition accuracy for complex actions like "ascending stairs" and "descending stairs" is relatively low. After adding the MHA mechanism to the SA-BiGRU, the proposed model was obtained. The recognition results are shown in Figure 9. At this time, the overall accuracy has been significantly improved. This is because the MHA has the advantage of capturing various aspects of the data simultaneously, which can extract more valuable information and improve the performance of the proposed model. From Table 4, it can be observed that the AUC metric also shows progressive improvement, particularly after the introduction of the MHA mechanism. MHA is capable of better capturing the multi-dimensional features of the data, significantly improving the AUC value. This indicates that the proposed model not only excels in overall classification accuracy but also demonstrates a notable advantage in its ability to distinguish between categories.

Table 4. Results of ablation study based on the PAMAP2 dataset

Model	ACC (%)	F1 (%)	AUC
GRU	90.03	89.95	0.921
Bi-GRU	90.84	90.79	0.928
SA-BiGRU	91.13	91.15	0.934
MHA-GRU	95.68	95.71	0.969
MHA-BiGRU	96.26	96.34	0.978
Proposed model	**98.06**	**98.13**	**0.982**

For the WCBD dataset, the comparison results of the proposed model against the baseline models can be found in Table 5. Compared with GRU, Bi-GRU can improve the accuracy of action recognition by 0.69%. MHA-BiGRU achieves an improvement of 0.66% in accuracy of action recognition compared to MHA-GRU. This is because Bi-GRU, as a bidirectional gated recurrent unit, has the capability of simultaneously processing forward and backward time-series information. In HAR, since the sensor signals of human actions have complex front-to-back correlations, traditional GRU that processes time-series data in only one direction may lose some important information, while Bi-GRU can fully capture it in two directions. For example, in climbing movements, the actions of the arms and legs are interrelated in time, and the features of actions in both forward and reverse time dimensions are obtained by Bi-GRU at the same time, thereby understanding the action pattern more comprehensively and improving the recognition accuracy. Compared

TABLE 5. Results of ablation study based on the WCBD dataset

Model	ACC (%)	F1 (%)	AUC
GRU	92.45	92.48	0.906
Bi-GRU	93.14	93.69	0.917
SA-BiGRU	93.85	93.91	0.929
MHA-GRU	94.77	94.79	0.936
MHA-BiGRU	95.43	95.39	0.942
Proposed model	**95.86**	**95.91**	**0.948**

with Bi-GRU, MHA-BiGRU achieves a 2.29% higher accuracy in action recognition. The experimental results show that the performance of Bi-GRU can be enhanced by the MHA mechanism. Since different action parts and time points have different importance for recognition, MHA is capable of automatically learning and attending to those parts and time points that are more critical for action recognition. Additionally, it can adaptively adjust weights, thereby highlighting important information and suppressing irrelevant information. For example, in climbing movements, the features of key actions such as hand grasping and leg pedaling will be more important at certain moments in the time-series. The MHA mechanism can capture this key information and make the model more focused when processing data, thereby improving the classification level and generalization ability of the model. Notably, compared with the MHA-BiGRU model, the proposed model has an accurate improvement of 0.43%. This is mainly because the SA mechanism can accurately focus on the key feature information in the action sequence, thus enhancing the performance of the model in the human action recognition task. Combining the multiple attention mechanisms with Bi-GRU can greatly improve the recognition of human actions, the experiment further illustrates the effectiveness of the proposed model. Meanwhile, with model optimization, the AUC is continuously improving, indicating that the model's ability to distinguish key information in climbing actions has significantly improved, leading to better classification confidence.

To verify the robustness of the model proposed, Gaussian white noise with varying variances was introduced to the sensor data in the WCBD datasets, the accuracy of the proposed model and other models was observed under the influence of different noise levels. The experimental results of this part are presented in Figure 11. As the noise enhances, the performance of the proposed model and the methods compared all exhibit a downward trend. Nevertheless, the performance decrease of the proposed model is significantly smaller than that of the other methods on the WCBD dataset. The better performance of the proposed model can be attributed to the multiple attention mechanisms. In climbing actions, when noise affects the acceleration data in a certain direction, some attention heads of the MHA can still extract valid features from the relatively less-affected acceleration components. The SA can ensure that the model can focus on the key feature information in the action sequence even when it is interfered by noise. In contrast, the baseline models lack comprehensive attention mechanisms. The GRU, which processes time-series data in only one direction, is more vulnerable to noise interference. Although the Bi-GRU can process bidirectional information, compared with the proposed model, its feature extraction ability is insufficient. When faced with noise, it is less effective in highlighting the key features. In summary, the combination of multiple attention mechanisms in the proposed model endows it with a stronger ability to handle noisy sensor data. As a result, in the noisy environment of the WCBD dataset, compared with other methods, the performance degradation of the model is smaller.

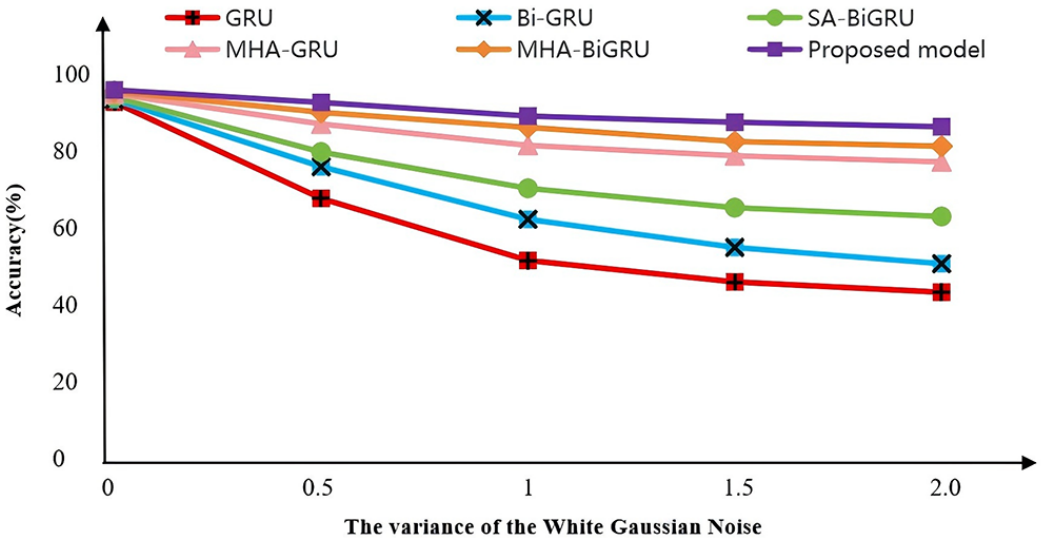


FIGURE 11. Accuracy of six models against different variances of noise on the WCBD dataset

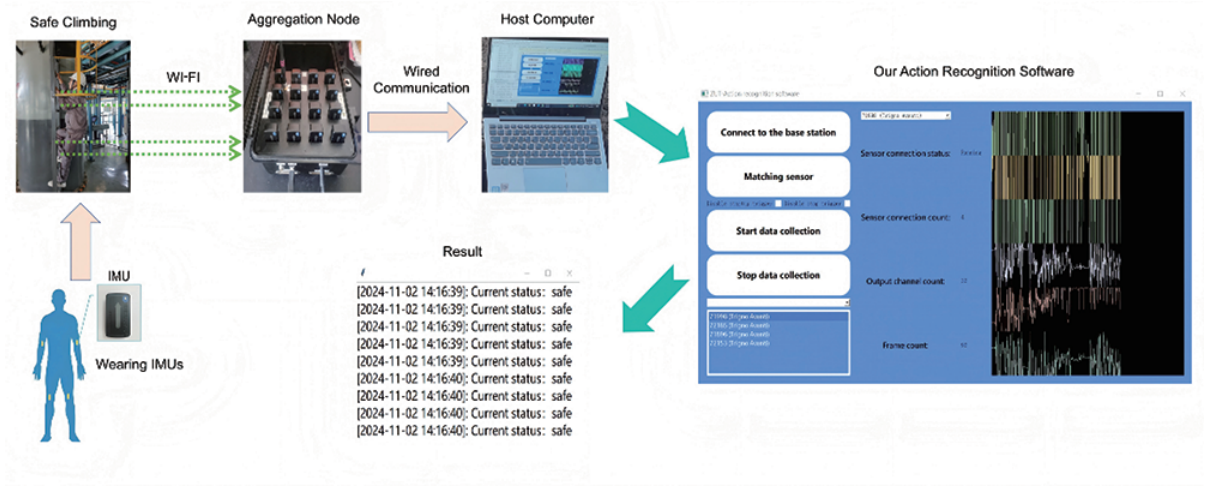


FIGURE 12. The deployment architecture of our HAR system in industrial environments

4. **Application.** As shown in Figure 12, to evaluate the practical performance of the model, the time-series signals collected by multiple IMUs are transmitted to the sensor aggregation node via Wi-Fi. Then, the IMU signals are transmitted to our host computer on which the proposed model is deployed. Finally, the action data is input into the model, and the results are displayed on the software, which can achieve the timely recognition of dangerous actions and effectively prevent accidents. During the recognition process, the software interface displays information such as the amount of data collected, data waveforms, and the status of the sensors. The proposed model is implemented on the TensorFlow 2.1 framework, with experiments conducted on a desktop computer equipped with an Intel Core i7-10700K CPU (3.8GHz, 8 cores), 32GB DDR4 RAM, and an NVIDIA RTX 3090 GPU. The system runs Ubuntu 20.04 LTS with CUDA 11.2. The average processing time for each set of sensor data is 0.2 seconds, which is significantly lower than the allowable delay time for action monitoring in industrial environments. This ensures real-time recognition of workers' unsafe behavior, thereby effectively preventing human-induced accidents.

5. **Conclusions.** To recognize the unsafety activities of workers in the industrial environment, a novel HAR method is proposed based on deep Bi-GRU with the multiple attention mechanism. The MHA mechanism is used to effectively obtain comprehensive information from multiple wearable sensors by linear transformation and multiple attention heads and the residual connection is designed to avoid gradient disappearance. In addition, a deep Bi-GRU with the SA module is introduced to highlight the important temporal features of human activities. Several experiments were conducted on both PAMAP2 and WCBD datasets and the results verified that the proposed model performs better performance and robustness than the existing state-of-the-art methods. Moreover, our HAR software system applied in the industrial environment has been developed with the proposed model and the model has demonstrated certain potential with the help of IoT frameworks. In future work, the proposed method will be further improved, and we will carry out research on knowledge distillation strategy to compress the model for its real-time application.

Acknowledgment. This work was supported by Henan Province Science and Technology R&D projects (242102320215), high-end foreign expert program of Henan province (HNGD2024032), Key scientific research project plan of colleges and universities in Henan Province (24A413013), Natural Science Foundation of Henan Province (No. 2423004214 17), the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (No. 2024C03254, 2025C04005), Key Research and Development Projects in Henan Province (No. 251111211 600) and Subject strength enhancement plan project of Zhongyuan University of Technology (GG202412).

REFERENCES

[1] X. Luo, Q. Liu and Z. Qiu, The influence of human-organizational factors on falling accidents from historical text data, *Frontiers in Public Health*, vol.9, 2022.

[2] A. M. Vukicevic, M. N. Petrovic, N. M. Knezevic and K. M. Jovanovic, Deep learning-based recognition of unsafe acts in manufacturing industry, *IEEE Access*, vol.11, pp.103406-103418, 2023.

[3] Y. Albadawi and T. Shanableh, Hand-crafted features with a simple deep learning architecture for sensor-based human activity recognition, *IEEE Sensors Journal*, vol.24, no.17, pp.28300-28313, DOI: 10.1109/JSEN.2024.3422272, 2024.

[4] V. S. Suh et al., Worker activity recognition in manufacturing line using near-body electric field, *IEEE Internet of Things Journal*, vol.11, no.7, pp.11554-11565, DOI: 10.1109/JIOT.2023.3330372, 2024.

[5] H. Han, G. Kim, S. Choi, A. Basu and S. W. Yoon, Human activity and correlated posture monitoring using earlobe-worn wearable sensor system and deep learning algorithm, *IEEE Sensors Journal*, vol.24, no.1, pp.533-542, DOI: 10.1109/JSEN.2023.3332897, 2024.

[6] S. Zhang, H. Zhou, R. Tchantchane and G. Alici, A wearable Human-Machine-Interface (HMI) system based on colocated EMG-pFMG sensing for hand gesture recognition, *IEEE/ASME Transactions on Mechatronics*, vol.30, no.1, pp.369-380, DOI: 10.1109/TMECH.2024.3386929, 2025.

[7] X. Yan, H. Li, A. R. Li and H. Zhang, Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention, *Automation in Construction*, vol.74, pp.2-11, 2017.

[8] H. Zhu and B. Hwang, Real-time safety and worker self-assessment: Sensor-based mobile system for critical unsafe behaviors, *Automation in Construction*, 2025.

[9] D. Lamani, P. Kumar, A. Bhagyalakshmi et al., SVM directed machine learning classifier for human action recognition network, *Scientific Reports*, vol.15, 672, 2025.

[10] T. F. N. Bukht, H. Rahman and A. Jalal, A novel framework for human action recognition based on features fusion and decision tree, *2023 4th International Conference on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, pp.1-6, DOI: 10.1109/ICACS55311.2023.10089752, 2023.

[11] N. A. Choudhury and B. Soni, An efficient and lightweight deep learning model for human activity recognition on raw sensor data in uncontrolled environment, *IEEE Sensors Journal*, vol.23, no.20, pp.25579-25586, DOI: 10.1109/JSEN.2023.3312478, 2023.

[12] B. Boufama, S. Hussein, E. Kim and I. Ahmad, A deep-learning approach for task recognition of industrial workers and RULA score calculation, *2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC)*, Marrakech, Morocco, pp.1-6, DOI: 10.1109/ISIVC 61350.2024.10577819, 2024.

[13] S. Mekruksavanich, P. Jantawong, N. Hnoohom and A. Jitpattanakul, Wearable-based activity recognition of construction workers using LSTM neural networks, *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Phuket, Thailand, pp.1-4, 2022.

[14] W. Li, X. Sun, T. He and T. Jiang, Development of a human activity recognition algorithm based on BiLSTM for construction workers, *2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE)*, Shanghai, China, pp.198-204, DOI: 10.1109/ICAACE 61206.2024.10548765, 2024.

[15] S. Yang, X. Yu and Y. Zhou, LSTM and GRU neural network performance comparison study: Taking yelp review dataset as an example, *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, Shanghai, China, pp.98-101, DOI: 10.1109/IWECAI 50956.2020.00027, 2020.

[16] H. Wu, Z. Huang, W. Jiang and X. Zhao, Facial expression recognition algorithm based on multi-attention mechanism, *International Journal of Innovative Computing, Information and Control*, vol.19, no.4, pp.1239-1250, 2023.

[17] X. Xie, C. Li, Y. Liu, J. Song, J. Ahn and Z. Zhang, Application of YOLOv4 algorithm with integrated attention mechanism in metal surface defect detection, *International Journal of Innovative Computing, Information and Control*, vol.19, no.2, pp.447-463, 2023.

[18] S. Mekruksavanich, P. Jantawong, W. Phaphan and A. Jitpattanakul, A sensor-based deep learning approach for recognizing daily and work activities in open environments for sanitation workers, *2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, Chiang-mai, Thailand, pp.577-581, DOI: 10.1109/ECTIDAMTNCON60 518.2024.10480100, 2024.

[19] H. Zhang and L. Xu, Multi-STMT: Multi-level network for human activity recognition based on wearable sensors, *IEEE Transactions on Instrumentation and Measurement*, vol.73, pp.1-12, Article no.2508612, 2024.

[20] Y. Wang et al., MhaGNN: A novel framework for wearable sensor-based human activity recognition combining multi-head attention and graph neural networks, *IEEE Transactions on Instrumentation and Measurement*, vol.72, pp.1-14, Article no.2514314, 2023.

[21] A. Reiss and D. Stricker, Introducing a new benchmarked dataset for activity monitoring, *Proc. of the 16th International Symposium on Wearable Computers*, pp.108-109, 2012.

[22] S. Wang et al., Robust human activity recognition via wearable sensors using dynamic Gaussian kernel learning, *IEEE Sensors Journal*, vol.24, no.6, pp.8265-8280, 2024.

[23] Y. Tang, L. Zhang, F. Min and J. He, Multiscale deep feature learning for human activity recognition using wearable sensors, *IEEE Transactions on Industrial Electronics*, vol.70, no.2, pp.2106-2116, DOI: 10.1109/TIE.2022.3161812, 2023.

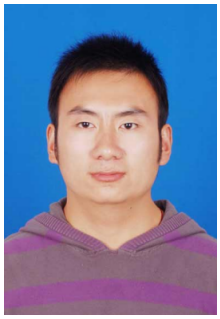
[24] M. Yao et al., Revisiting large-kernel CNN design via structural re-parameterization for sensor-based human activity recognition, *IEEE Sensors Journal*, vol.24, no.8, pp.12863-12876, DOI: 10.1109/ JSEN.2024.3371462, 2024.

[25] L. Lu and T. Deng, A method of self-supervised denoising and classification for sensor-based human activity recognition, *IEEE Sensors Journal*, vol.23, no.22, pp.27997-28011, 2023.

Author Biography

Haiquan Wang received his B.Eng. and Ph.D. degrees from Northwestern Polytechnical University, China, in 2006 and 2009, respectively. He is currently a professor at the Zhongyuan-Pertersburg Aviation College, Zhongyuan University of Technology. His research interests include intelligent control and complex system health management.

Jiabo Zhai received a Bachelor's degree in Engineering from Zhongyuan University of Technology, China, in 2022. He is currently a postgraduate student at Zhongyuan University of Technology, China, Grade 2022. His main research interests include deep learning and human activity recognition.



Yueyi Yang received his Ph.D. degree from Beijing Jiaotong University, China, in 2022. He is currently a lecturer at Zhongyuan-Pertersburg Aviation College, Zhongyuan University of Technology. His research interests include train communication networks, fault diagnosis, and health management.

Xiaobin Xu received the Ph.D. degree in Power Electronics and Power Transmission from Shanghai Maritime University, Shanghai, China, in 2009. He is currently a professor with the Department of Automation and The Belt and Road Information Research Institute, Hangzhou Dianzi University, Hangzhou, China. His research interests include evidence theory, fuzzy set theory and applications in the processing of uncertain information.

Shengjun Wen is currently a professor with Zhongyuan-Petersburg Aviation College in Zhongyuan University of Technology. He received the Ph.D. degree in Electronic and Information Engineering from Graduate School of Engineering from Tokyo University of Agriculture and Technology in 2011. His research interests include nonlinear control, safety control and robotics.

Jinxia Wen received a Bachelor's degree in Automation from Henan University of Technology, China, in 2022. She is currently a graduate student of Zhongyuan University of Technology. Her main research interests include federated learning and human activity recognition.

Yabo Hu received a Bachelor's degree in Electrical Engineering and Automation from Zhongyuan University of Technology, China, in 2024. He is currently a post-graduate student in Zhongyuan University of Technology. His main research interests are deep learning and image processing.