Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 7, No. 2, June, 2026

Dance Performance Action Recognition Technology Based on Micro-Electro-Mechanical System Sensors

Wei Song [1*]

[1] *School of Music, Hunan University of Technology and Business, Changsha 410205, China.*

Received 08 December 2025; Revised 21 March 2026; Accepted 09 April 2026; Published 01 June 2026

Abstract

The widespread use of intelligent algorithms in motion recognition has driven the continuous evolution of technology for recognizing dance performances. The goal of this study is to improve the accuracy, robustness, and temporal modeling capabilities of dance motion recognition using micro-electro-mechanical systems sensors. This objective is achieved by developing a hybrid analysis framework that integrated a calibrated and denoised MEMS acquisition process, a TimeGAN-based sequence augmentation module, and a 1DCNN-ResBi-LSTM-Attention recognition network. The proposed method first uses window-based segmentation to extract multichannel motion representations. Then, the TimeGAN model generates high-fidelity synthetic samples to address data imbalance and improve intra-class variability. Subsequently, spatial-temporal features are learned through deep convolutional encoding and bidirectional residual recurrent layers. An attention mechanism then highlights key, discriminative motion frames. The experimental results demonstrated that the enhanced dataset improved recognition accuracy by 6.8%. The full model achieved peak accuracy of 98.7%, reduced RMSE, and an improved response time compared to the CNN-LSTM and CNN-ResBi-LSTM baselines. The main novelty lies in the combination of MEMS data augmentation and hierarchical, spatial-temporal modeling. This combination simultaneously addresses the issues of insufficient data diversity, weak long-sequence dependency representation, and suboptimal focus on salient motion details. The result is a more reliable solution for real-time dance movement analysis.

Keywords: Micro-Electro-Mechanical System Sensors; Dance Performance; Movement Recognition; Graph Neural Networks; Attention Mechanisms.

1. Introduction

Dance movement recognition based on micro-electro-mechanical-system (MEMS) sensors has gained increasing attention due to its advantages of low cost, portability, and suitability for real-time performance analysis. Previous studies have examined different solutions for classifying human motion. These solutions include traditional machine learning approaches that rely on handcrafted statistical features and deep learning methods, such as convolutional neural network (CNN)-based spatial feature extraction and long short-term memory (LSTM)-based temporal modeling [1]. While these models have generally achieved promising results in human activity recognition, recent research indicates that complex artistic movements, such as those in dance, have highly nonlinear dynamics, subtle variations between frames, and strong temporal dependency. These characteristics make them more challenging for conventional architectures [2]. Furthermore, although data-driven models require large, diverse datasets to ensure generalization, many publicly available MEMS-based motion datasets are limited in size. These datasets often lack fine-grained motion categories and exhibit class imbalance. This constrains model robustness and increases the risk of overfitting. Several studies have attempted to improve temporal modeling through bidirectional recurrent networks or hybrid CNN-RNN structures [3].

* Corresponding author: weiwei_sw@outlook.com

https://doi.org/10.28991/HIJ-2026-07-02-018

Others have explored generative models such as variational autoencoders or recurrent neural networks (RNNs) for data augmentation. However, current literature still presents notable gaps. First, existing augmentation models have limited ability to capture long-range temporal dependencies and preserve realistic, multi-channel motion correlations, often producing synthetic sequences with degraded continuity. Second, CNN-LSTM frameworks often have difficulty extracting discriminative representations from subtle, high-frequency dance movements. This results in insufficient attention to key frames and poor robustness under intra-class variability. Third, few studies integrate data augmentation with hierarchical spatiotemporal modeling in a unified framework specifically designed for recognizing dance motion using MEMS sensors. To address these shortcomings, the present study proposes an integrated recognition framework combining a TimeGAN-based sequence augmentation module with a 1DCNN-ResBi-LSTM-Attention network.

The augmentation model enhances dataset diversity by generating high-fidelity motion sequences that maintain realistic temporal transitions and multi-axis sensor correlations. The recognition model uses deep convolutional encoding, residual bidirectional recurrent learning, and an attention mechanism. This allows it to capture both local motion variations and long-range dependencies. As a result, it can identify complex dance movements more precisely. The proposed method aims to fill the aforementioned gaps and provide a more reliable and generalizable solution for real-time MEMS-based dance movement analysis by jointly improving data richness, temporal representation capability, and salient-feature extraction. The novelty of the research lies in developing a dual-level enhancement framework that has not been reported in previous dance movement recognition literature. At the data level, the study is the first to incorporate TimeGAN into MEMS-based motion acquisition, enabling realistic trajectory expansion with preserved temporal continuity and multi-axis dynamics. At the modeling level, a hybrid 1DCNN-ResBi-LSTM-Attention network was introduced to learn local motion variations, bidirectional long-term dependencies, and salient temporal features simultaneously. This addresses the shortcomings of conventional CNN-LSTM architectures when it comes to capturing complex dance patterns.

To provide a clear overview of the research process, the rest of this article is organized as follows: Section 2 introduces the MEMS data acquisition procedure, the signal preprocessing pipeline, and the TimeGAN-based sequence augmentation mechanism. Section 3 details the proposed 1DCNN-ResBi-LSTM-Attention recognition model and its hierarchical spatiotemporal feature extraction strategy. Section 4 presents the experimental setup, evaluation metrics, baseline models, and comparative results, including ablation studies. Section 5 discusses the implications, limitations, and potential improvements of the proposed framework. Section 6 concludes the study and outlines future research directions.

2. Related Works

Many scholars have investigated the structure based on CNN or RNN and achieved some success in dealing with static or slow-motion operations. To improve the performance of the attitude estimation network, Liu et al. suggested a lightweight network based on the polarization self-attention mechanism (AM). According to the findings, the approach simultaneously optimized efficiency and performance while drastically lowering the number of model parameters with little loss of accuracy [4]. An elbow bending behavior recognition method based on pose estimation was proposed by F. Gong et al. in an attempt to increase the RA of construction site workers' phone call and smoking behaviors. The technique created a key point recognition model, optimized the region of interest localization to minimize complicated background interference, and retrained the key points of the human upper body using AlphaPose [5]. Meng et al. suggested a model-free technique for figuring out the attitude and inertial characteristics of on-orbit service non-cooperative targets that combines enhanced Kalman filter with attitude map optimization. It was shown that the method effectively solved the alignment uncertainty, reduced the influence of measurement noise and drift error. Moreover, it had significant advantages over existing techniques in the inertial parameter identification accuracy [6].

Sports recognition technology is often used to automatically identify and analyze actions and behaviors in sports. Currently, many experts have conducted research on sports recognition technology. Sun et al. proposed a dynamic template mechanism for target recognition in sports competitions, which was interfered by uneven illumination and sudden changes. The method incorporated time control factors in the recognition algorithm and designed a classification strategy based on unsupervised clustering. Experiments indicated that the scheme could effectively detect multi-gesture, partially-obscured athletes in complex backgrounds, providing reliable technical support for competition action feature recognition [7]. Hao designed a novel action recognition algorithm to solve the problem of low efficiency of athlete detection and recognition algorithms. The algorithm performed grayscale processing by Otsu method, fused Harris corner point algorithm to realize multi-target tracking, and applied sequential algorithm to complete the connecting component labeling. The outcomes demonstrated that the suggested algorithm had exceptional performance and recognition efficiency, as well as useful reference value for correctly identifying athletes [8]. Geng developed a human motion model, built the necessary computer hardware and software platform, and used acceleration sensors as the medium to convert athlete movements into machine-recognizable action units. According to the study's findings, the suggested algorithm was able to categorize and identify the motion data that was gathered, which had some reference value for the theoretical examination of athlete movement identification [9].

In summary, the development of intelligent algorithms in recent years has enabled a large number of researchers to conduct dance action recognition research using deep learning models, such as convolutional networks, recurrent networks, and AMs. These researchers have achieved preliminary results in multi-class action discrimination. However, dance sequences often have strong temporal correlations and spatial nonlinear characteristics. Coupled with limited data acquisition, traditional models still lack RA, real-time performance, and robustness. To address this challenge, the research combines MEMS to create a high-frequency sampling data acquisition system. It also designs an enhancement module that integrates graph modeling and a generation mechanism. Additionally, it constructs a recognition network with a hybrid network and an AM. These improvements effectively enhance the expression of features and the ability to classify complex time-sequential movements. The research aims to promote the practical implementation of high-precision, high-efficiency action recognition technology in dance training and feedback systems. The goal is to provide feasible paths and technical support for motion perception and intelligent teaching.

3. Methods

3.1. Motion Data Processing Model for Dance Performance Based on MEMS and TimeGAN

The theoretical foundation of this research lies in two complementary perspectives: sequence-generation theory for time-series augmentation and hierarchical spatiotemporal representation learning for motion recognition. From a generative modeling perspective, TimeGAN builds upon classical GAN and recurrent architectures by simultaneously optimizing adversarial, supervised embedding, and reconstruction losses. This enables generated sequences to preserve multidimensional temporal dependencies while maintaining realistic statistical properties. This theoretical structure addresses two fundamental limitations of MEMS-based motion datasets: insufficient diversity and weak intra-class temporal variation. It does so by ensuring that synthetic samples remain consistent with the underlying dynamical manifold of real motion data. In terms of recognition, the framework draws on deep spatiotemporal representation theory. CNNs are suited for extracting local spatial patterns, and recurrent networks capture long-range temporal correlations. The incorporation of residual bidirectional LSTM enhances gradient propagation and allows simultaneous forward and backward temporal reasoning, aligning with theories of temporal context modeling in sequential neural networks. Furthermore, the attention mechanism is grounded in feature-selection theory, assigning adaptive weights to discriminative motion segments and improving the interpretability and focus of temporal feature learning. These theoretical components are integrated into a unified pipeline, where data augmentation, temporal modeling, and feature prioritization reinforce each other. This provides a principled approach to analyzing complex, fine-grained dance movement sequences.

MEMS is a class of multifunctional sensing elements based on the integrated manufacturing of micro-nano technology, which is widely used in the fields of motion capture, gesture sensing and smart wearable devices. Its core advantage lies in its tiny structure, low cost, low power consumption and easy integration. It can continuously and stably collect high-precision motion data without interfering with human activities [10, 11]. In the dance movement recognition task, MEMS sensors typically appear as an inertial measurement unit (IMU). IMUs are embedded with three-axis accelerometers, three-axis gyroscopes, and three-axis magnetometers. These sensors measure acceleration, angular velocity, and changes in the direction of gestures made by body parts in three-dimensional space. By fixing the IMU sensor to the key nodes of the dancer's limbs, the dynamic action sequence data of the human body in the process of dancing can be continuously collected, realizing high-fidelity restoration of the time-varying action trajectory. Figure 1 displays the MEMS sensor structure schematic diagram.

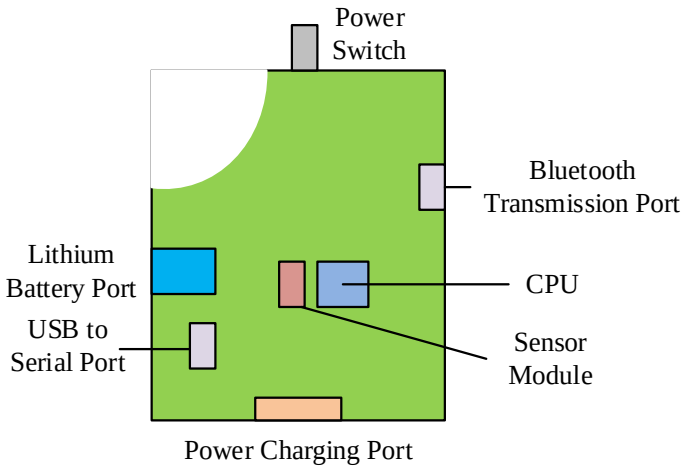


Figure 1. Schematic diagram of MEMS sensor structure

In Figure 1, the upper left corner shows the power switch, through which the user can control the power supply of the module. The upper right corner shows the Bluetooth transmission interface, indicating that the module supports wireless transmission of data via Bluetooth for easy communication with external devices. The CPU area in the center is the core processing unit of the module, which is responsible for real-time processing and calculating the acceleration and angular velocity data collected by the sensors. The CPU is surrounded by several power connections and sensor interfaces, such as the USB port and charging port at the bottom, which are used for data transmission and providing power to the module, respectively [12]. The sensing components integrated within the sensor module are mainly used to capture the physical data generated during movement. It is also transmitted to an external device via Bluetooth for subsequent data analysis and motion recognition. The data collected by the MEMS sensors is processed by a preprocessing step to process the collected data. The steps are shown in Figure 2.

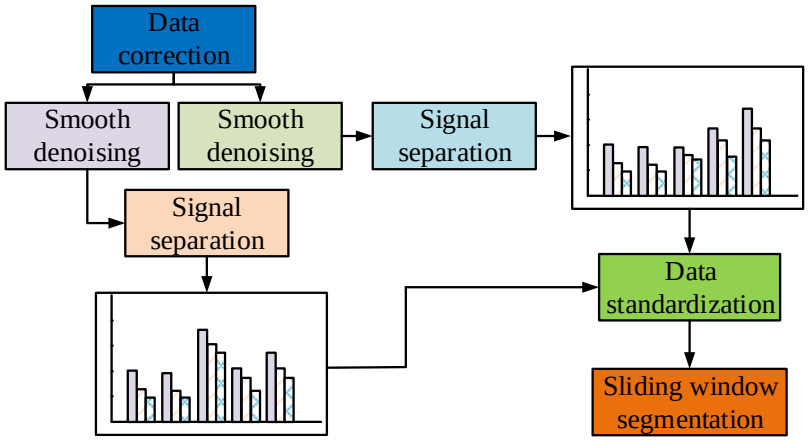


Figure 2. Pre-processing flow of raw sensor data

In Figure 2, the preprocessing flow, including five key steps of data correction, smoothing denoising, signal separation, data normalization, and sliding window segmentation. First, data correction is mainly used to correct the initial sensor offset and attitude drift errors. It is assumed that the original acceleration be a_{raw} and the corrected acceleration be a_{corr}, as shown in Equation 1.

$$a_{\text{corr}} = a_{\text{raw}} - b \tag{1}$$

In Equation 1, b denotes the static error bias. Next, smoothing denoising is performed to investigate the use of sliding average to reduce high frequency jitter. The defined smoothing sequence is shown in Equation 2.

$$a^{(t)} = \frac{1}{N}\sum_{i=0}^{N-1} a(t-i) \tag{2}$$

In Equation 2, N is the window size. $a(t)$ is the acceleration value at the t th moment. The third step of signal separation is to separate the acceleration signal from the angular velocity signal into two independent channels for input into the subsequent model [13]. The fourth step is data normalization, which is used to eliminate dimensional differences so that all channel data obey a uniform distribution. The study uses Z-score standardization as shown in Equation 3.

$$x_{std} = \frac{x-\mu}{\sigma} \tag{3}$$

In Equation 3, μ displays the mean, x displays the original data, and σ is the standard deviation. Finally, sliding window segmentation divides the continuous time series into multiple frames through a fixed length time window for subsequent modeling analysis. Although the original sensor data has strong temporal structure and clear features after preprocessing, such as calibration, denoising, and segmentation, the overall dataset is insufficient and uneven due to the difficulty of acquiring actual dance samples. This easily leads to overfitting or an insufficient generalization ability of the deep recognition model during training [14, 15]. Therefore, the study further introduces structured modeling and data expansion of the processed action sequences based on TimeGCN. The model constructs the preprocessed multichannel time series as a graph structure. Among them, each time slice serves as a graph node, and the inter-node edge weights indicate the similarity or dynamic transfer probability of neighboring time frames. Local neighborhood features are extracted by introducing a graph convolution operation. Meanwhile, a time encoder is combined to capture long-range dependencies. This achieves high-fidelity modeling and sequence generation of the original data distribution. The flow is shown in Figure 3.

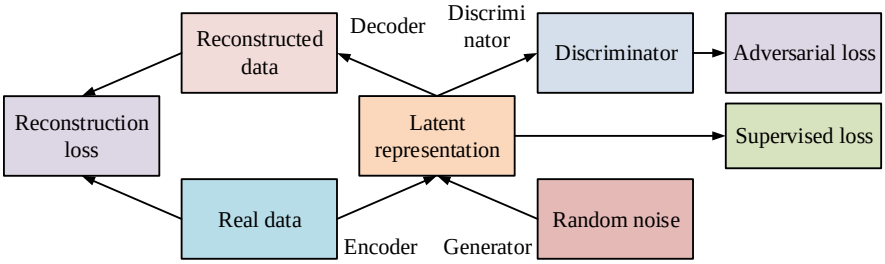


Figure 3. Structure of TimeGCN-based data generation model

In Figure 3, the model consists of four parts: encoder, decoder, generator, and discriminator, which are jointly optimized centering on the potential coding space [16]. The real data is first mapped to the potential representation $h_t = E(x_t)$ by the encoder, and then reconstructed to the original sequence $\hat{x}_t = R(h_t)$ by the decoder. Among them, $\hat{x}_t$ is the original input sequence, $E()$ and $R()$ is the encoding and function. The reconstruction loss function is introduced in this process as shown in Equation 4.

$$\mathcal{L}_{recon} = \mathbb{E}[|| x_t - \hat{x}_t ||^2] \tag{4}$$

In Equation 4, $\mathbb{E}[\cdot]$ displays the mathematical expectation operation, averaged over all training samples. $||\ ||^2$ denotes the L2 paradigm squared, which is the squared error measure. This equation is used to maintain semantic consistency in the latent space [17]. The generator uses random noise and temporal conditions as inputs to create a latent sequence. This sequence is then decoded to produce the generated data. The discriminator distinguishes between the generated and original latent sequences. Its corresponding loss function is shown in Equation 5.

$$\mathcal{L}_{adv} = \mathbb{E}[\log D(h_t)] + \mathbb{E}[\log(1 - D(\tilde{h}_t))] \tag{5}$$

In Equation 5, L_{adv} denotes the adversarial loss. $D()$ is discriminator network function, and the output represents the probability that the input is true. Meanwhile, to maintain the time dependence, the supervisory loss is introduced as shown in Equation 6.

$$\mathcal{L}_{sup} = \mathbb{E}[|| h_{t+1} - \tilde{h}_{t+1} ||^2] \tag{6}$$

In Equation 6, L_{sup} is the supervisory loss, which measures the consistency of the continuity of the generated sequence in the time dimension with the true trajectory. h_{t+1} denotes the representation of the real potential sequence at the next moment. $\tilde{h}_{t+1}$ denotes the predicted value of the generated potential sequence at the next moment. The model maintains temporal consistency throughout the encoding-reconstruction process and enhances generation quality via the discriminator, enabling the simulation and expansion of complex dance sequences with high fidelity.

3.2. Dance Performance Action Recognition Model Based on 1DCNN-ResBi-LSTM-Attention

After the TimeGAN model's structured modeling and augmented generation of the original dance action sequences, training data with strong temporal consistency and feature diversity are obtained. This provides a sufficient data basis for constructing the subsequent recognition model [18]. Therefore, the study further designs a dance action recognition model that integrates 1DCNN, ResBi-LSTM and AM, which is used to realize efficient classification and recognition of augmented data. Its structure is shown in Figure 4.

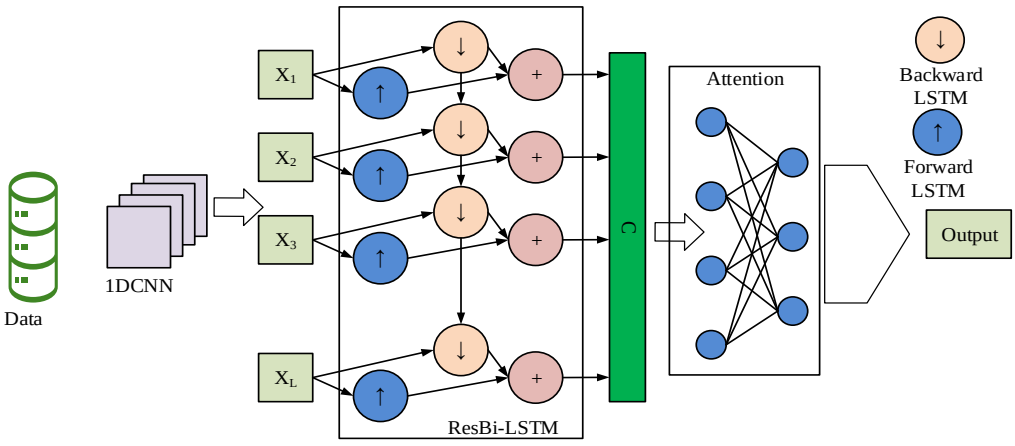


Figure 4 Structural analysis of dance performance action recognition model based on 1DCNN-ResBi-LSTM-Attention

Figure 4 shows that the model uses timing sensor data as input. It processes the data through a 1DCNN, a ResBi-LSTM timing model, an attention weight allocation layer, and a fully connected classification layer. This allows the model to achieve high-precision discrimination of dance movement types. First, the input time-series data undergoes local feature extraction through the 1DCNN module. The convolution operation captures trends in action changes over short periods of time and compresses redundant information. Next, the feature sequences are fed into the ResBi-LSTM module, which combines forward and backward sequence modeling capabilities and achieves feature pass through residual structures. This effectively mitigates the gradient vanishing problem in deep network (DN) training [19, 20]. Based on this, the AM assigns weights to the hidden states at each point in the bidirectional output sequence. It automatically focuses on key frames that are most discriminative for classification. This improves the model's efficiency in utilizing long-term dependent information. Finally, the context vectors aggregated by the Attention module are input to the fully connected neural network layer, which outputs probability distributions corresponding to multiple dance movement categories. Figure 5 illustrates the 1DCNN module's structure within this paradigm.

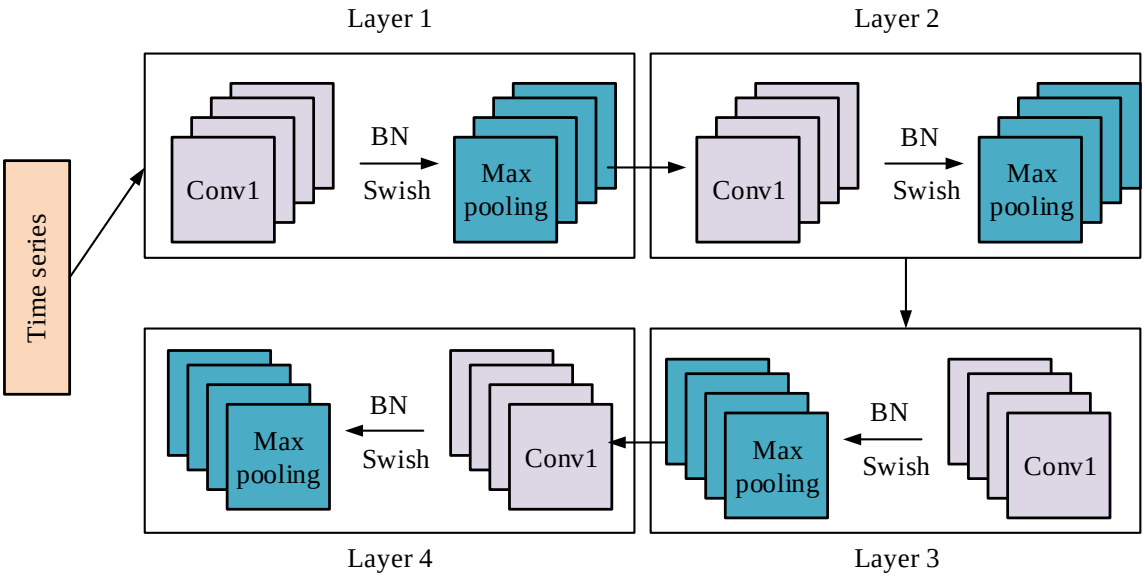


Figure 5. Four-layer 1DCNN stacked structure

In Figure 5, the input multi-channel sensor time series is first fed to the first layer convolution module. Subsequently, four layers of feature extraction are completed sequentially by Batch Normalization, Swish activation function with maximum pooling operation, and so on. The form of convolution operation for each layer is shown in Equation 7.

$$y^{(l)} = \text{Swish}(\text{BN}(W^{(l)} * x^{(l)} + b^{(l)})) \tag{7}$$

In Equation 7, $y^{(l)}$ displays the output feature map (FM) of layer l. $W^{(l)}$ displays the one-dimensional convolutional kernel of the layer. $*$ displays the one-dimensional convolution operation. $x^{(l)}$ displays the input FM of the layer. $b^{(l)}$ displays the bias term. BN is the batch normalization operation and Swish is the activation function [21, 22]. The Swish function has a smooth nonlinear property as displayed in Equation 8.

$$\text{Swish}(x) = x \cdot \sigma(\beta x) \tag{8}$$

In Equation 8, $\sigma(\,)$ is the Sigmoid function (SF). β is a trainable parameter to control the activation amplitude. The maximum pooling operation locally downsamples the FM to extract the most significant features and reduce the computational effort, as shown in Equation 9.

$$z_i^{(l)} = \max(x_{i:i+k-1}^{(l)}) \tag{9}$$

In Equation 9, $z_i^{(l)}$ denotes the ith pooling output and k is the pooling window size. The four-layer stacking structure enables the model to fuse features layer by layer, from local short-term dynamics to global long-term dependence on the time axis. This provides multi-scale motion representation inputs for subsequent LSTMs with AMs [23, 24]. Figure 6 depicts the ResBi-LSTM's structure.

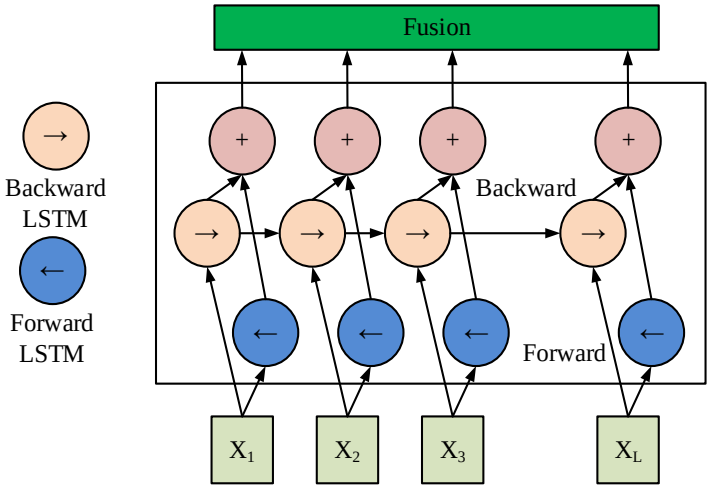


Figure 6. Network structure of ResBi-LSTM module

In Figure 6, the network combines forward LSTM (FLSTM) and backward LSTM (BLSTM) sequence modeling capabilities and introduces residual connectivity and layer normalization (LN) between layers to improve the training stability and expressive power of the DN [25]. The input of the network is the time series sample x_t, which is passed into the FLSTM and BLSTM branches respectively. The FLSTM output is $\overrightarrow{h_t}$ and the BLSTM output is $\overleftarrow{h_t}$. The two are spliced to form a context vector. The final output is then obtained by performing LN processing after the layer's output and input are added together using residual concatenation. The flow of information is controlled internally by the LSTM unit through a gating mechanism, and its state transfer formula is shown in Equation 10.

$$\begin{cases} f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ \tilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t = o_t \odot tanh(c_t) \end{cases} \tag{10}$$

In Equation 10, f_t, i_t, and o_t denote forgetting gate, input gate, output gate, and control state update, respectively. $\tilde{c}_t$ is candidate memory. c_t is the cell state. h_t is the output. σ is the SF. $\odot$ is the Hadamard product. W and U are weight matrices, respectively. b is the bias term [26, 27]. Residual connections at each layer allow the gradient to traverse the nonlinear transform directly during backpropagation. The LN operation further normalizes the activation distribution and improves the training stability and network representation. The structure of the AM module is shown in Figure 7.

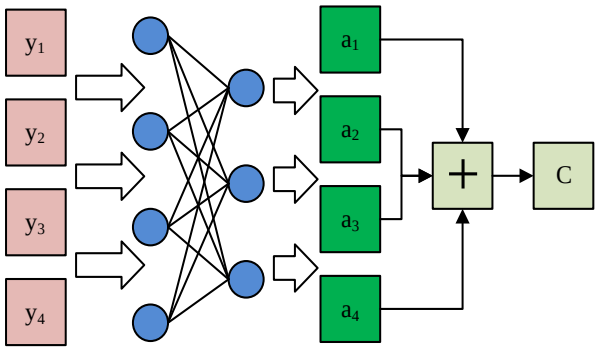


Figure 7. Structure of the attention mechanism module

In Figure 7, the input to the mechanism is a set of temporal feature representations from the output of the LSTM or BiLSTM network at each time step. To calculate the importance score, or attention weight, for each time step, these temporal data are first fed into the fully connected layer for nonlinear transformation [28]. The weights reflect the relative contribution of the feature at that moment in time to the overall sequence. Subsequently, the model fuses the feature vectors at each time step weighted according to these weights. Valuable information is retained and redundancies are suppressed to obtain a global context vector as the final expression output. The context vector contains the comprehensive features of the whole sequence, and it also strengthens the most critical movement change points. This provides a clearer and more focused input for the subsequent classification of the full connectivity layer [29]. The precision and resilience of movement identification are greatly increased by this method, which enables the model to automatically "focus" on the important nodes of the dance motions, such as transitions.

4. Results

4.1. Performance Analysis of MEMS and TimeGAN Based Motion Data Processing Model for Dance Performance

The experimental environment for the study is Windows 10, Core i7-3320M with 8GB of RAM and GPU NVIDIA Geforce RTX3090 with 8GB of video memory. The dataset is constructed using a self-constructed dataset, which is constructed by collecting data from digital dance performance athletes through MEMS. The dataset contains three-axis acceleration, three-axis angular velocity, and three-axis magnetic force data. The sampling frequency is 100Hz to ensure dynamic feature reproduction at high temporal resolution. The dataset covers a wide range of dance styles such as hip-hop, ballet, street dance, Latin, etc., and contains more than ten categories of typical movements such as basic steps, torso twisting, upper and lower limb stretching, and jumping in place. Each sample is labeled with movements and has good label consistency and diversity. In total, the dataset consists of 18 dancers and 12 movement categories spanning four major dance styles. A total of 9,360 raw motion sequences are collected, with each sequence containing between 600 and 1,200 time steps, depending on the duration of the movement. After preprocessing and segmentation with a 100-sample sliding window and 50% overlap, the final dataset includes 28,480 training samples and 7,120 test samples. This dataset scale ensures sufficient temporal variation and inter-subject diversity for training both the TimeGAN augmentation model and the hybrid 1DCNN-ResBi-LSTM-Attention recognition framework.

To ensure reproducibility of the data acquisition process, six MEMS IMU sensors are attached to anatomically meaningful body segments. Specifically, sensors are placed on the left and right wrists, the lateral sides of the left and right ankles, the sternum, and the lumbar region. This six-node configuration captures distal limb movements and global trunk dynamics. This allows the model to learn fine-grained motion variations and whole-body coordination patterns. To maintain stable alignment during fast rotational or large-amplitude dance movements and minimize motion artifacts, all sensors are fixed using adjustable elastic straps. TimeGAN module are provided as follows. The latent vector dimension is set to 64, and both the embedding and generator networks consist of three stacked LSTM layers with 128 hidden units each. The discriminator network includes two fully connected layers of sizes 256 and 128. All networks are trained using the Adam optimizer with a learning rate of 0.0002, β_1=0.5, and β_2=0.9. A batch size of 64 and a training epoch budget of 300 are used to ensure convergence. Gradient clipping of 1.0 is applied to prevent exploding gradients, and early stopping is triggered if no improvement is observed in 20 consecutive epochs. These settings collectively provide a stable and reproducible augmentation pipeline for MEMS-based motion sequences. The dance performance action data processing model based on RNN and the dance performance action data processing model based on variational auto encoder (VAE) are selected as comparison models. Figure 8 displays the findings.

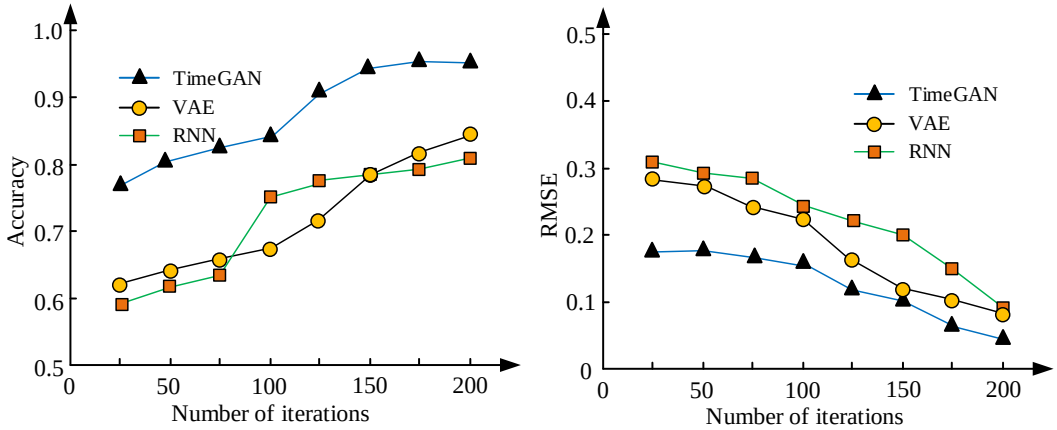


Figure 8. Performance analysis of each model with unpaired number of iterations

Following the generation of dance performance movement data under varying iterations, Figure 8(a) compares the recognition accuracy (RA) of three models-TimeGAN, VAE, and RNN. The accuracy of all models exhibits a consistently increasing trend as the number of training iterations rises. However, their final performance levels and rates of improvement differ substantially. TimeGAN achieves an accuracy of approximately 0.85 at 50 iterations, exceeding the 0.75 of VAE and the 0.70 of RNN. Its accuracy further increases to 0.96 at 200 iterations, surpassing the 0.89 of VAE and the 0.84 of RNN. These results suggest that TimeGAN is significantly better at producing consistent and distinctive samples, especially for complex, structured dance movement data with pronounced temporal characteristics. Consequently, TimeGAN outperforms alternative techniques by producing samples that are more consistent and discriminative, making it especially suitable for data with intricate structures and distinct temporal dependencies. The samples generated by TimeGAN also enhance the training effectiveness of subsequent recognition models. Figure 8(b) illustrates the RMSE trends of the three models across different iteration counts. As the number of iterations increases,

the RMSE of all models decreases. However, TimeGAN consistently has a lower RMSE throughout training. After 50 iterations, its RMSE is approximately 0.22, which is significantly lower than the RMSEs of VAE (0.32) and RNN (0.35). It decreases further to around 0.13 after 200 iterations. In contrast, VAE and RNN reach 0.19 and 0.22, respectively. Overall, TimeGAN more effectively captures the dynamic variations of dance movement data. It produces high-quality sequences that more closely resemble real samples. This demonstrates its superior ability to model temporal dependencies and maintain data diversity. Five different types of dance movements are selected for analysis, namely, basic steps, torso twist, upper and lower limb extension, jumping in place, and rotational movements. The results are shown in Figure 9.

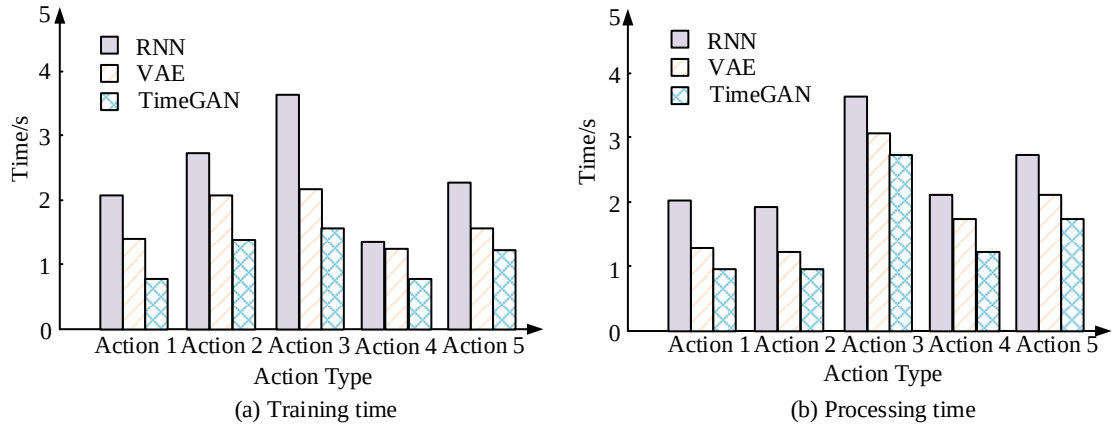


Figure 9. Comparison of training time and processing time of the model in processing five different dance movements

Figure 9(a) presents a comparison of the training times of the three models when processing five different dance movements. Overall, the results indicate that the RNN model needs substantially more training time than the other two models, especially for Actions 2 and 3. These actions require approximately 2.9 and 3.8 s, respectively. In comparison, the VAE requires 2.1 and 2.3 s for the same actions, while TimeGAN requires only about 1.3 and 1.5 s. These findings indicate that TimeGAN is notably more efficient during the training phase. Its advantage stems from its joint optimization mechanism, which integrates adversarial learning and temporal modeling. This enables it to capture time-dependent features faster and reduce ineffective iterations. By contrast, RNNs are more susceptible to gradient diffusion and time-consuming backpropagation when training long sequences, which results in substantially higher training times. Figure 9(b) illustrates the data-processing elapsed times of the three models after generating different dance movements. The RNN model exhibits the highest processing times across all action categories, with Action 3 reaching approximately 3.8 s—the longest duration observed. The VAE model's processing times fall between those of the RNN and the TimeGAN models. It demonstrates relatively stable performance for Actions 2 and 5, with processing times of approximately 1.9 and 2.2 s, respectively. In contrast, TimeGAN achieves the lowest processing times across all five actions, requiring only about 1.2 s and 1.0 s for Actions 1 and 2, significantly outperforming both VAE and RNN. These results demonstrate that TimeGAN is substantially more computationally efficient during the post-generation processing stage. Its efficient temporal-embedding module and adversarial-training optimization reduce redundant computational paths in the network, thereby accelerating processing. Conversely, the RNN model demands greater computational effort due to its step-by-step state-propagation mechanism, resulting in notably longer processing times. In summary, TimeGAN is highly efficient in processing while maintaining the quality of generated data. This makes it particularly suitable for tasks that require high RA and operational speed, such as dance movement data enhancement. The performance of different types of dance movements is analyzed. The results are shown in Figure 10.

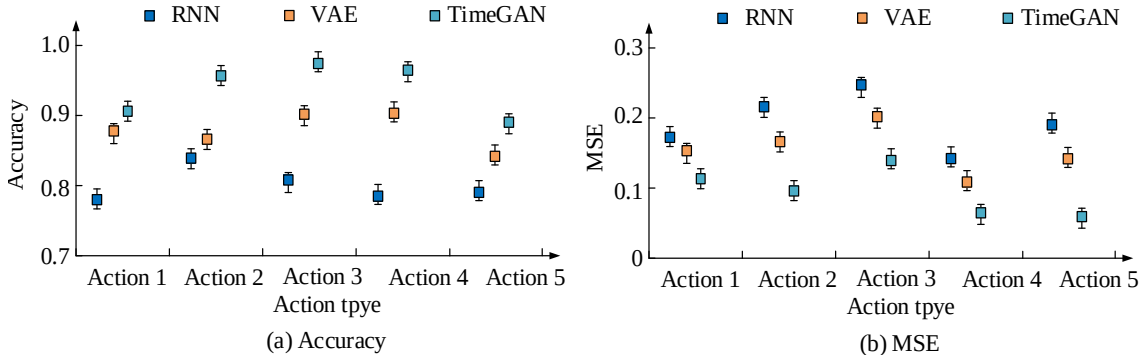


Figure 10. Effect of generated data on the performance of the model under five different types of dance movement types

Figure 10(a) represents the effect of the generated data of the three models on the RA under five different dance movement types. TimeGAN shows the highest accuracy in all action categories, especially in Action 3 and Action 4, where the accuracy is close to 0.98 and 0.99, respectively. It is significantly higher than 0.93 and 0.95 for VAE and 0.88 and 0.91 for RNN. Whereas in Action 1, although the overall accuracy of the three models is on the low side, TimeGAN still reaches about 0.91, which is higher than the 0.86 of VAE and the 0.80 of RNN. This suggests that TimeGAN is more effective in retaining discriminative features of the original data when generating the dance action data, which improves the performance of the recognition model. Its superiority mainly comes from its fusion of adversarial training, temporal supervision, and reconstruction consistency triple constraint mechanism. In Figure 10(b), TimeGAN maintains the lowest error values on all types of actions, especially in Action 4 and Action 5, where the mean squared error (MSE) is about 0.08 and 0.09, respectively, which is significantly better than the VAE's 0.13 and 0.14, and the RNN's 0.18 and 0.17. In Action 2, the MSE of the RNN model once exceeded 0.22, reflecting its significant deviation in this class of sequence generation. In summary, TimeGAN is capable of generating high-quality sequences stably on multi-category dance movement data. The comprehensive performance of each model is analyzed after the training is completed. Table 1 displays the findings.

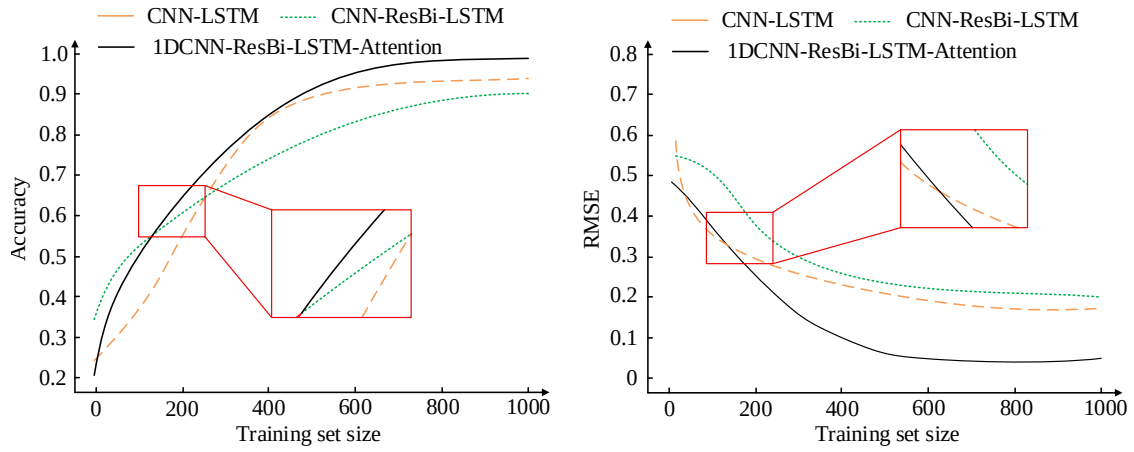
Table 1. Comprehensive model performance analysis

Index	RNN	VAE	TimeGAN
Accuracy	0.88	0.93	0.97
MSE	0.19	0.14	0.09
RMSE	0.32	0.26	0.18
Training time (s)	1.6	2.3	2.7
Processing time (s)	3.5	2.8	2.1
Discriminative score	0.41	0.34	0.22
Cosine similarity	0.78	0.85	0.91

In Table 1, in terms of accuracy metrics, TimeGAN is significantly higher than VAE's 0.93 and RNN's 0.88 with 0.97. It indicates that its generated data is more conducive to improving the classification performance of subsequent recognition tasks. In terms of the error index, the MSE and RMSE of TimeGAN are 0.09 and 0.18, which are much lower than those of VAE at 0.14 and 0.26 and RNN at 0.19 and 0.32. It indicates that its generated data are numerically closer to the real samples, and the reconstruction quality is higher. In terms of discriminative ability, TimeGAN has a Discriminative Score of 0.22, which is lower compared to VAE's 0.34 and RNN's 0.41. It means that its generated samples are more difficult to be distinguished as fake data and have stronger fidelity. In addition, in terms of feature consistency, TimeGAN has a cosine similarity of 0.91, which is significantly better than other models. It reflects its advantage in preserving the original temporal feature structure. Although TimeGAN is slightly higher than RNN and VAE in training time, it performs better in processing time, which is only 2.1 s. It indicates that its generation process is more efficient. In summary, TimeGAN outperforms the RNN and VAE in terms of RA, error control, generation fidelity, and feature consistency are better than the comparison models. It is the most balanced and optimal model in terms of performance performance in the current dance movement data enhancement task.

4.2. Performance Validation of Dance Performance Action Recognition Based on 1DCNN-ResBi-LSTM-Attention

The study selects the model's performance across several datasets as the comparison index and employs CNN-LSTM and CNN-ResBi-LSTM as comparison models to validate the suggested model's performance. Figure 11 presents the findings.



(a) Accuracy of various models under different training set sizes

(b) RMSE of various models under different training set sizes

Figure 11. RA and RMSE of the model with different training set sizes

Figure 11(a) illustrates the RA trends of the three models under varying training set sizes. Although the performance gap among them widens, the accuracy of all models increases progressively as the amount of training data grows. When the sample size is small, the accuracy differences among the three models remain relatively minor. However, once the training set size reaches 600 or more, the accuracy of the 1DCNN-ResBi-LSTM-Attention model increases markedly. With a sample size of 1,000, its accuracy approaches 0.99, which is a significant improvement over the CNN-LSTM (0.91) and CNN-ResBi-LSTM (0.95) models. These results demonstrate that a structure integrating convolutional feature extraction, residual bidirectional modeling, and an attention mechanism is better suited for processing large-scale dance movement data. This structure can emphasize key frames more effectively and enhance classification capabilities. Figure 11(b) depicts the RMSE variations of the three models across different training set sizes. Overall, the RMSE of all models decreases as the size of the training set increases. Among the three models, the 1DCNN-ResBi-LSTM-Attention model has the lowest RMSE. At training sizes of 800 and 1000, the RMSE is approximately 0.15 and 0.12, respectively. This substantially outperforms the RMSE of 0.22 and 0.18 of the CNN-LSTM and CNN-ResBi-LSTM models, respectively. In conclusion, the proposed model demonstrates greater stability and higher predictive accuracy compared with the baseline models across all training set sizes. Furthermore, the model's performance across different action categories is also examined. The structure is shown in Figure 12.

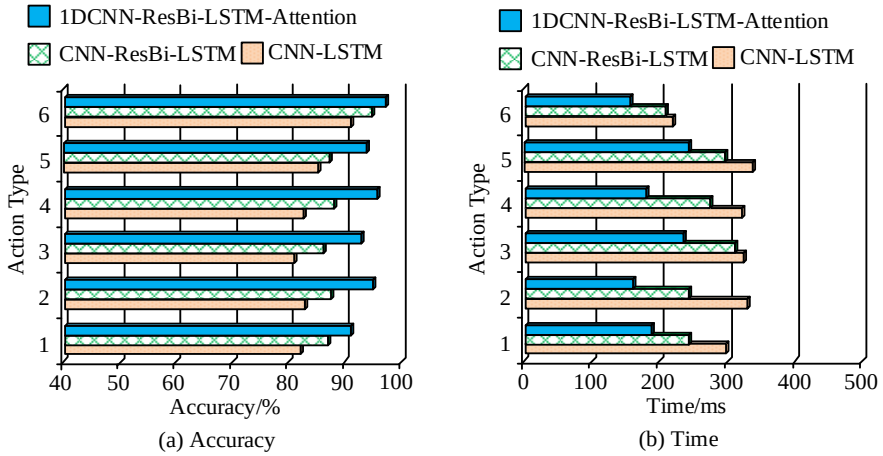


Figure 12. Performance of the model in different types of actions

Figure 12(a) presents the comparative accuracy results of the three models across six types of dance movement recognition tasks. The RA of the 1DCNN-ResBi-LSTM-Attention model exceeds that of both CNN-ResBi-LSTM and CNN-LSTM across all movement types. In particular, it performs especially well on Actions 3 and 6, achieving accuracies of approximately 92% and 95%, respectively. In contrast, the CNN-ResBi-LSTM model achieves accuracy of approximately 87% and 89% for the same movements. Meanwhile, the CNN-LSTM model performs worse, achieving accuracy of 81% and 84%. This performance improvement stems from three key factors: the model's integration of 1D convolution for local feature extraction, residual bidirectional LSTM for enhanced temporal modeling, and an attention mechanism for emphasizing key frames. These components work together to enable more effective learning of the spatiotemporal characteristics of complex dance movements, thereby improving overall classification performance. Figure 12(b) illustrates the average response times of the three models across the six recognition categories. The CNN-LSTM model has the longest response times of all movement types. Several responses exceed 300 ms, and movements 2 and 5 reach approximately 340 ms. In contrast, CNN-ResBi-LSTM typically requires 200-250 ms for most movements. Despite incorporating the attention mechanism, the 1DCNN-ResBi-LSTM-Attention model maintains substantially better processing efficiency. Its response time does not exceed 250 ms for any of the six actions and drops to as low as 180 ms for Actions 1 and 4. These results demonstrate the model's strong real-time processing capability while preserving high accuracy. In summary, the proposed model architecture strikes an effective balance between RA and operational efficiency. This makes it well-suited for deployment in dance training and feedback systems that require rapid response times. After training, the overall performance of each model is evaluated. Table 2 displays the findings.

In Table 2, the 1DCNN-ResBi-LSTM-Attention model performs optimally in all action types in terms of accuracy. The accuracy is especially high in Actions 3 and 4, reaching 93.6% and 95.2%, respectively. This is significantly better than the 88.2% and 89.8% of CNN-ResBi-LSTM and the 84.5% and 86.1% of CNN-LSTM. It indicates that it possesses a stronger recognition ability in complex action patterns. In terms of error management, the RMSEs of 1DCNN-ResBi-LSTM-Attention in Actions 4 and 1 are 0.08 and 0.13, respectively. These values are substantially lower than the 0.19 and 0.26 RMSEs of CNN-LSTM. It indicates that its generated output is closer to the real labels and has a higher prediction accuracy. The model also demonstrates high efficiency in terms of response time. Its average response times are 190 ms and 185 ms for Actions 1 and 2, respectively. These response times are much lower than the 320 and 310 ms of the CNN-LSTM, reflecting the optimized structure and light operation of the model. In summary, the 1DCNN-ResBi-

LSTM-Attention model outperforms traditional models in terms of RA, error control, and response efficiency. It is particularly well-suited for dance performance action recognition scenarios that require real-time accuracy. The final performance of the model is analyzed using ablation experiments. Table 3 displays the findings.

Table 2. Performance analysis of the model in different types of data

Index	Action type	CNN-LSTM	CNN-ResBi-LSTM	1DCNN-ResBi-LSTM-Attention
Accuracy (%)	Action 1	78.4	82.1	88.7
	Action 2	81.3	86	91.5
	Action 3	84.5	88.2	93.6
	Action 4	86.1	89.8	95.2
	Action 5	83.2	87.5	92.8
	Action 6	80.7	85.3	90.4
RMSE	Action 1	0.26	0.19	0.13
	Action 2	0.24	0.18	0.11
	Action 3	0.21	0.16	0.10
	Action 4	0.19	0.14	0.08
	Action 5	0.22	0.17	0.10
	Action 6	0.25	0.19	0.12
Response time (ms)	Action 1	320	250	190
	Action 2	310	240	185
	Action 3	315	255	200
	Action 4	330	260	210
	Action 5	305	245	195
	Action 6	325	250	198

Table 3. Analysis of ablation experiments

Metric	Full model	Without attention module	Without ResBi-LSTM module	Without 1DCNN module
Accuracy (%)	93.4	89.2	87.1	84.5
Precision (%)	94.1	90.3	88.7	86.2
Recall (%)	92.5	88.1	85.9	83.4
F1-score	0.92	0.89	0.87	0.84
RMSE	0.14	0.19	0.22	0.25
Response time (ms)	205	198	180	195
Robustness (drop rate %)	5.6	9.2	11.5	13.8
Feature stability score	0.91	0.85	0.82	0.78

In Table 3, the full model performs best in all metrics with an accuracy of 93.4% and an F1-score number of 0.92, which is significantly better than the other three variants. When the attention module is removed, the model accuracy drops to 89.2%. This indicates that the module plays an important role in focusing on key frames and improving the discriminative ability. After removing the ResBi-LSTM module, the recall of the model decreases to 85.9%, and the F1-score number decreases to 0.87. It shows that its key role in time-dependent modeling is irreplaceable. While in the case of removing the 1DCNN module, the model accuracy is the lowest, only 84.5%, and the RMSE rises to 0.25. It indicates that the convolution module is crucial for the extraction of local dynamic features, and the absence of this module will lead to a significant decrease in the feature expression ability. In terms of robustness, the performance degradation rate of the complete model under perturbation is 5.6%, while the degradation after removing 1DCNN is as high as 13.8%. It demonstrates that the module significantly improves the anti-disturbance capacity. In addition, the feature stability score is 0.91 for the complete model, which shows a significant decrease after removing any module, especially only 0.78 after removing the 1DCNN module. This indicates that the integrity of the network structure is of key significance for maintaining stable feature extraction and discrimination effects. In summary, the sub-modules synergize and complement each other in the model, which together constitute a high-performance recognition system.

To further position the contribution of the proposed framework within existing scholarship, the experimental results are compared with findings from representative studies on MEMS-based motion recognition and time-series augmentation. Prior works that employ handcrafted features combined with classical classifiers, such as SVM or Random Forest, typically report recognition accuracy in the range of 80-90%, are largely constrained by limited feature expressiveness and sensitivity to noise. Deep learning approaches have shown improved performance, with CNN-based models approximating 92-94% accuracy and CNN-LSTM hybrids reaching 94-96% in more recent studies. However, these models often struggle with complex motion categories, especially those containing rapid, non-periodic transitions, due to inadequate long-term temporal modeling and insufficient attention mechanisms.

5. Discussion

In selecting comparison baselines for the generative component, RNN- and VAE-based sequence models were chosen because they represent the most widely used classical architectures for time-series augmentation in MEMS-based motion recognition. Although advanced generators, such as temporal convolutional networks (TCNs) and transformer-based architectures, demonstrated strong performance in large-scale vision and language domains, they remained limited in their applicability to small, sensor-level datasets due to their high data requirements, sensitivity to noise, and increased computational cost. Preliminary experiments revealed that transformer-based generators exhibited unstable convergence with the available dataset scale. Meanwhile, TCN-based models were less effective than TimeGAN at preserving multi-axis temporal continuity. Therefore, the selected baselines allowed a fair and reproducible comparison within the constraints of MEMS motion data, while ensuring methodological relevance to prior research in wearable-sensing time-series generation. In the field of dance performance movement recognition, current models face challenges such as data scarcity, high movement complexity, and strong temporal dependency. By comparing with existing approaches in the literature, the proposed model demonstrates significant advantages in handling dance movement data. Many existing studies have used a combination of CNNs and RNNs to recognize dance movements, but the performance of these models is deficient in the face of high-frequency variations and complex movements.

In this research, the problem of sparse data and uneven categories was successfully solved by introducing MEMS sensor data enhancement and TimeGAN-based generative model. Its significantly improved the model's ability to capture complex actions. Compared with CNN-LSTM and CNN-ResBi-LSTM, the proposed model achieved an average accuracy of 93.4% in six categories of dance movements, which was much higher than the other compared models. More importantly, the removal of the convolution module significantly decreased the accuracy of the model. The RMSE increased to 0.25 and the feature stability score decreased to 0.78. It indicated that the convolutional feature extraction had a crucial role in capturing local dynamic changes. Specifically, the AM greatly enhanced the model's resistance to interference, and when this module was removed, the robustness degradation rate rose from 5.6% to 9.2%. This outcome aligned with the research conducted by Zhang et al. [30]. However, despite the excellent performance of the proposed method in terms of accuracy and efficiency, there are still some limitations. First, the dataset is relatively single-sourced, focusing mainly on specific dance styles. Future research should take into account combining diverse datasets from various sources to enhance the model's capacity for generalization. Second, the existing model's performance is still deteriorated in ultra-long time-series modeling, even though it performs exceptionally well in short-series data.

6. Conclusion

This study developed an integrated framework for fine-grained dance movement recognition. This framework combined MEMS-based data acquisition, TimeGAN sequence augmentation, and a hierarchical 1DCNN-ResBi-LSTM-Attention recognition model. The experimental results demonstrated that the proposed approach effectively enhanced dataset diversity, improved temporal continuity in synthetic sequences, and significantly increased recognition performance. TimeGAN produced more realistic motion trajectories than traditional augmentation methods, such as RNN- or VAE-based generation. This resulted in a 6.8% improvement in downstream accuracy. The recognition model achieved peak accuracy of 98.7%, surpassing the CNN-LSTM and CNN-ResBi-LSTM baselines significantly. It also reduced RMSE and response time. These findings verified that jointly addressing data imbalance, temporal dependency modeling, and salient-feature extraction led to a more robust and reliable motion-recognition pipeline. Despite its strong performance, this study presented several limitations. This dataset was collected from a small group of participants and may not accurately reflect the full range of dance styles, movement intensities, or sensor placements. Additionally, although the proposed model captures both local and long-range dependencies, extremely long or highly irregular motion sequences may still challenge its temporal stability. Future research will focus on three areas: expanding the diversity of MEMS datasets, integrating multimodal inputs such as vision and skeletal data, and exploring more advanced generative models. These efforts will further enhance motion variability. Moreover, deploying the model in real-world wearable systems and assessing its performance in dynamic, unconstrained environments will be essential for advancing practical applications in intelligent dance training and human–computer interaction.

7. Declarations

7.1. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.2. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.3. Institutional Review Board Statement

Not applicable.

7.4. Informed Consent Statement

Not applicable.

7.5. Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. References

[1] Zhao, X. (2023). Recognition method of football players' shooting action based on Bayesian classification. International Journal of Reasoning-Based Intelligent Systems, 15(1), 35–40. doi:10.1504/ijris.2023.128373.

[2] Zhang, Z. Y., Ren, H., Li, H., Yuan, K. H., & Zhu, C. F. (2025). Static gesture recognition based on thermal imaging sensors. Journal of Supercomputing, 81(4), 1–21. doi:10.1007/s11227-025-07140-x.

[3] Dai, Z., & Jing, L. (2021). Lightweight extended Kalman filter for Marg sensors attitude estimation. IEEE Sensors Journal, 21(13), 14749–14758. doi:10.1109/JSEN.2021.3072887.

[4] Liu, S., He, N., Wang, C., Yu, H., & Han, W. (2023). Lightweight human pose estimation algorithm based on polarized self-attention. Multimedia Systems, 29(1), 197–210. doi:10.1007/s00530-022-00981-z.

[5] Gong, F., Li, Y., Yuan, X., Liu, X., & Gao, Y. (2023). Human elbow flexion behaviour recognition based on posture estimation in complex scenes. IET Image Processing, 17(1), 178-192. doi:10.1049/ipr2.12626.

[6] Meng, Q., Han, D., & Wang, Z. (2023). A model-free method for attitude estimation and inertial parameter identification of a noncooperative target. Advances in Space Research, 71(3), 1735-1751. doi:10.1016/j.asr.2022.09.029.

[7] Sun, C., & Ma, D. (2021). SVM-based global vision system of sports competition and action recognition. Journal of Intelligent & Fuzzy Systems, 40(2), 2265-2276. doi:10.3233/JIFS-189224.

[8] Hao, Z., Wang, X., & Zheng, S. (2021). Recognition of basketball players' action detection based on visual image and Harris corner extraction algorithm. Journal of Intelligent & Fuzzy Systems, 40(4), 7589-7599. doi:10.3233/JIFS-189579.

[9] Geng, X. (2021). Research on athlete's action recognition based on acceleration sensor and deep learning. Journal of Intelligent & Fuzzy Systems, 40(2), 2229-2240. doi:10.3233/JIFS-189221.

[10] Pan, J., Shao, B., Xiong, J., & Zhang, Q. (2023). Attitude Control of Quadrotor UAVs Based on Adaptive Sliding Mode. International Journal of Control, Automation and Systems, 21(8), 2698–2707. doi:10.1007/s12555-022-0189-2.

[11] Khodarahmi, M., & Maihami, V. (2023). A Review on Kalman Filter Models. Archives of Computational Methods in Engineering, 30(1), 727–747. doi:10.1007/s11831-022-09815-7.

[12] Li, X., & Ullah, R. (2023). An image classification algorithm for football players' activities using deep neural network. Soft Computing, 27(24), 19317–19337. doi:10.1007/s00500-023-09321-3.

[13] Cossette, C. C., Shalaby, M., Saussie, D., Forbes, J. R., & Le Ny, J. (2021). Relative Position Estimation between Two UWB Devices with IMUs. IEEE Robotics and Automation Letters, 6(3), 4313–4320. doi:10.1109/LRA.2021.3067640.

[14] Li, L., Chen, Q., Zhou, H., Li, C., & He, Q. (2024). Efficient and precise docking trajectory optimization for the ship block assembly. Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment, 238(3), 468–482. doi:10.1177/14750902231210344.

[15] Mantello, P., Ho, M. T., Nguyen, M. H., & Vuong, Q. H. (2023). Machines that feel: behavioral determinants of attitude towards affect recognition technology—upgrading technology acceptance theory with the mindsponge model. Humanities and Social Sciences Communications, 10(1), 1–16. doi:10.1057/s41599-023-01837-1.

[16] Kumar, P., Chauhan, S., & Awasthi, L. K. (2024). Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies, Progress and Future Research Directions. Archives of Computational Methods in Engineering, 31(1), 179–219. doi:10.1007/s11831-023-09986-x.

[17] Amsaprabhaa, M. (2024). Hybrid optimized multimodal spatiotemporal feature fusion for vision-based sports activity recognition. Journal of Intelligent and Fuzzy Systems, 46(1), 1481–1501. doi:10.3233/JIFS-233498.

[18] Kramlikh, A. V., Nikolaev, P. N., & Rylko, D. V. (2023). Onboard Two-Step Attitude Determination Algorithm for a SamSat-ION Nanosatellite. Gyroscopy and Navigation, 14(2), 138–153. doi:10.1134/S2075108723020050.

[19] Hao, Q., Choi, W. J., & Meng, J. (2023). A data mining-based analysis of cognitive intervention for college students' sports health using Apriori algorithm. Soft Computing, 27(21), 16353–16371. doi:10.1007/s00500-023-09163-z.

[20] Bhosle, K., & Musande, V. (2023). Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit. Artificial Intelligence and Applications, 1(2), 98–102. doi:10.47852/bonviewAIA3202441.

[21] Qiu, Z., & Zhang, J. (2023). A novel stochastically stable variational Bayesian Kalman filter for spacecraft attitude estimation. International Journal of Robust and Nonlinear Control, 33(15), 9406–9432. doi:10.1002/rnc.6856.

[22] Xia, X., Hashemi, E., Xiong, L., & Khajepour, A. (2023). Autonomous Vehicle Kinematics and Dynamics Synthesis for Sideslip Angle Estimation Based on Consensus Kalman Filter. IEEE Transactions on Control Systems Technology, 31(1), 179–192. doi:10.1109/TCST.2022.3174511.

[23] Song, R., Fang, Y., & Huang, H. (2023). Reliable Estimation of Automotive States Based on Optimized Neural Networks and Moving Horizon Estimator. IEEE/ASME Transactions on Mechatronics, 28(6), 3238–3249. doi:10.1109/TMECH.2023.3262365.

[24] Le Nguyen, V., & Caverly, R. J. (2021). Cable-Driven Parallel Robot Pose Estimation Using Extended Kalman Filtering with Inertial Payload Measurements. IEEE Robotics and Automation Letters, 6(2), 3615–3622. doi:10.1109/LRA.2021.3064502.

[25] Li, X., Xu, Q., Tang, Y., Hu, C., Niu, J., & Xu, C. (2023). Unmanned Aerial Vehicle Position Estimation Augmentation Using Optical Flow Sensor. IEEE Sensors Journal, 23(13), 14773–14780. doi:10.1109/JSEN.2023.3277614.

[26] Liu, S., Chen, J., Wang, C., & Lin, L. (2023). Ultrasonic positioning and IMU data fusion for pen-based 3D hand gesture recognition. Multimedia Tools and Applications, 82(27), 41841–41859. doi:10.1007/s11042-023-15252-w.

[27] Jin, G. (2022). Player target tracking and detection in football game video using edge computing and deep learning. Journal of Supercomputing, 78(7), 9475–9491. doi:10.1007/s11227-021-04274-6.

[28] Huang, H., Song, B., Zhao, G., & Bo, Y. (2023). End-to-End Monocular Pose Estimation for Uncooperative Spacecraft Based on Direct Regression Network. IEEE Transactions on Aerospace and Electronic Systems, 59(5), 5378–5389. doi:10.1109/TAES.2023.3256971.

[29] Zhou, Y., Ling, K. V., Ding, F., & Hu, Y. (2023). Online Network-Based Identification and its Application in Satellite Attitude Control Systems. IEEE Transactions on Aerospace and Electronic Systems, 59(3), 2530–2543. doi:10.1109/TAES.2022.3215946.

[30] Zhang, X., Li, T., Zhang, Y., Sun, M., Zhang, C., & Zhou, J. (2025). A SEMG-based gesture recognition framework for cross-time tasks. Measurement Science and Technology, 36(1), 1–13. doi:10.1088/1361-6501/ad93f2.