IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 May 2025, accepted 9 July 2025, date of publication 18 July 2025, date of current version 25 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3590407

RESEARCH ARTICLE

Temporal and Modality Awareness-Based Lightweight Residual Network With Attention Mechanism for Human Activity Recognition Using a Lower-Limb Exoskeleton Robot

CHANG-SIK SON AND WON-SEOK KANG
Division of Intelligent Robot, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, South Korea

Corresponding author: Chang-Sik Son (changsikson@dgist.ac.kr)

This work was supported in part by the DGIST R&D Program of the Ministry of Science and ICT under Grant 25-IT-02; and in part by the Artificial Inteligence Learning Data Construction Support Project, the Ministry of Science and ICT under Grant 2022060038.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB), Gyeongsang National University Hospital, South Korea, under Application No. GNUCH 2022-08-007-001, and performed in line with the Declaration of Helsinki.

ABSTRACT Although many human activity recognition (HAR) models have achieved high accuracy, their computational complexity often limits deployment in systems with constrained hardware resources, such as wearable lower-limb exoskeletons. In addition, existing models frequently overlook the complementary nature of multimodal sensor signals, focusing primarily on temporal dynamics while underutilizing modality-specific information. To address these issues, this study proposes a lightweight residual network for recognizing diverse locomotion activities across varying terrains using multimodal sensing data. The model adopts an asymmetric convolutional architecture composed of depthwise and pointwise layers to efficiently capture temporal and modality-specific features while significantly reducing the number of trainable parameters. A channel-attention block is further integrated to emphasize salient fused features. Evaluations on the Walking Assist Wearable Robot Motion dataset, which contains kinematic and postural signals from 500 adults using a lower-limb exoskeleton, demonstrated that the proposed model achieves an accuracy of 98.23% and a macro F1 score of 98.21%, with only 48,037 parameters. This outperforms four hybrid deep learning baselines while reducing the parameter count by 5.5–12.9 times. To assess generalizability, additional experiments were conducted on four benchmark datasets—UCI-HAR, HAPT, PAMAP2, and WISDM—under varying batch size conditions. The proposed model consistently achieved competitive or superior macro F1 scores of 0.9627 ± 0.0046, 0.776 ± 0.0138, 0.8937 ± 0.0121, and 0.9736 ± 0.0025, respectively, confirming its robustness and adaptability across diverse real-world HAR scenarios.

INDEX TERMS Attention mechanism, depthwise separable convolution, human activity recognition, lower-limb exoskeleton robot.

I. INTRODUCTION

Wearable exoskeleton robots are intelligent mechatronics developed by combining the mechanism of the human body and mechanical properties to provide rehabilitation, motion assistance, and augmentation functions [1] to individuals with neuromuscular impairments to help them overcome mobility limitations and reacquire physical independence [2]. The robots are used for robot-assisted rehabilitation to help improve the physical functions of patients after a stroke [2], [3], [4] or vehicular accident [5], [6]. Recently, with the rapid growth of the population of older adults, indicating a general trend of an aging population, the risk of accidents such as falls has increased. Accordingly, wearable exoskeleton robots

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

are increasingly being used as augmentation exoskeletons to reduce the risk of accidents and assist the physical functions of elderly people [7], [8], [9].

The ultimate goals for the design and development of wearable exoskeleton robots have two dimensions: (1) understanding human intent and providing assistance as needed and (2) implementing a human–robot collaboration in which humans are dominant and robots are secondary [1]. To this end, motion-intensity recognition (MIR) and adaptive control are essential technologies that enable these robots to detect changes in the environment and transition between different locomotion modes. In particular, MIR allows real-time detection of the user's movements and serves as an essential precondition for active compliance control, providing stable walking assistance across different types of terrains. Therefore, accurate and efficient MIR will contribute to exoskeleton robots achieving natural human–robot collaboration [10], [11] and improving wearing comfort for wearers [12].

To realize such natural human–robot interactions, a number of studies have been conducted on the development of human motion-intention mechanisms that utilize bioelectric signals, such as electroencephalography (EEG) and electromyography (EMG) biosignals. Among the different types of biosignal measurement methods, EMG allows the detection of motor intention before the actual initiation of joint motions through surface EMG (sEMG) signals generated from muscle contractions [13] and is widely used in myoelectric prosthetics or the control of exoskeletons [14]. A key step in the MIR process is feature extraction, wherein hand-crafted features are extracted, typically from the time domain (e.g., mean absolute value, difference of mean absolute value, and root mean square), frequency domain (e.g., peak frequency, median frequency, and mean power frequency), or time–frequency domain (e.g., Fourier transform and wavelet transform). However, performing feature extraction on EMG signals is not necessarily straightforward, particularly when sEMG signals are involved, because although EMG signals contain local or global information, the feature extraction of multichannel sEMG signals is complicated by feature redundancy and the time-consuming process of feature extraction itself [15]. To address these problems, dimensionality reduction algorithms such as independent component analysis (ICA) or principal component analysis (PCA) have been adopted to reduce the dimensions of sEMG signals. Alternatively, deep learning techniques have gained interest in the fields of human MIR and human activity recognition because these methods enable end-to-end learning without involving complex feature engineering. In particular, in the design and development of exoskeleton robots, deep learning methods are applied in the classification of gait events [16], [17], kinematics motion parameter prediction [18], [19], and kinetics motion parameter prediction [20], [21].

Recent studies have investigated various approaches for improving motion intention and motion recognition performance via the characteristics of bioelectric signals (e.g., EMG) or biomechanical signals (e.g., IMU). Zhu et al. [14] developed a deep neural network (DNN)-based motion-intention model for the identification of different motions and transition modes for a soft lower limb exoskeleton robot connected to IMUs and a load cell. Narayan et al. [22] presented a novel design of a convolutional neural network (CNN) classifier that performs hierarchical classification for lower limb locomotion modes via IMUs. In addition, Son and Kang [23] developed a CNN model based on multimodal biosignals (specifically, EMG signals, hip angle/velocity, and posture) for the identification of different locomotion activities across different types of terrains. Although several studies have proposed effective motion-intention mechanisms and locomotion activity recognition methods, many of these approaches require computational resources that may exceed the constraints of embedded systems typically used in wearable exoskeleton robots. To achieve accurate detection of the user's motion intention and activities under such limited hardware conditions, the development of a lightweight model is essential.

This study presents a novel lightweight residual network (ResNet) architecture that leverages temporal dynamics and sensor modality-specific features for robust recognition of diverse human locomotion behaviors across various terrain types. The proposed architecture is rigorously evaluated against four hybrid deep learning methods—CNNGRU1, CNNGRU2, CNNResBiLSTM, and CNNBiGRU—to demonstrate its performance and efficiency. The main contributions of this work are as follows:

1. A lightweight ResNet is designed for human locomotion recognition on heterogeneous terrains, including level ground, stairs, and ramps, using sensor data acquired from rotary encoders and an inertial measurement unit (IMU) during the operation of a wearable lower-limb exoskeleton robot.
2. The proposed network integrates computationally efficient residual blocks, composed of asymmetrically designed depthwise and pointwise convolutions, to effectively extract temporal patterns and modality-specific representations from multimodal sensor data.
3. We conducted extensive ablation studies to systematically analyze the impact of various design choices on model performance and efficiency. These include architectural configurations (e.g., number of residual blocks, presence of residual connections, use of attention mechanisms, and activation function types), data-related settings (e.g., sensor modality selection and window size), and training parameters (e.g., batch size variations).
4. To verify the generalizability of our approach, we evaluated the proposed model on four publicly available benchmark datasets—UCI-HAR, HAPT, PAMAP2, and WISDM—by comparing its performance with that of hybrid deep learning methods, demonstrating

competitive or superior results across diverse human activity recognition tasks.

The remainder of this paper is organized as follows. Section II presents related work regarding human motion intention and activity recognition via wearable sensors. Section III introduces ResNet, which is based on learning temporal and modality features, including the attention mechanism. Section IV describes the experimental setup and comparison methods, presents the results of the ablation studies, and reports the performance of the proposed model on four publicly available benchmark datasets. Finally, Section V concludes the paper by presenting the conclusions of the study and suggestions for future research.

II. RELATED WORK

In this section, we present major research studies on the recognition of lower limb motion intention and locomotion behaviors of human users based on combining multisensor signals from a wearable exoskeleton robot. We also introduce key models based on deep learning that can be applied in motion-intention recognition and human activity recognition.

A. HUMAN LOCOMOTION RECOGNITION

Lower limb motions consist of sequences of continuously varying locomotion. Therefore, the recognition and control of continuous locomotion in lower-limb exoskeleton robots is more complex than discrete motion recognition; however, it has more value for practical application. Hu et al. [24] reported that in the control of a powered leg prosthesis for amputees, the fusion of neuromechanical signals (e.g., EMG, goniometer, and IMU signals) of the bilateral lower limbs significantly improved the accuracy of the control system for five locomotion activities: level walking (LW), ramp ascent (RA), ramp descent (RD), stair ascent (SA), and stair descent (SD), and eight transition modes. Offline analysis was performed for one above-knee amputee as a proof of concept for the general application of the approach, and all the classification results of the linear discriminant analysis (LDA) model with bilateral signals achieved notably lower error rates (overall: 1.43, steady-state: 0.76, transitional: 4.5) than those obtained via the support vector machine and ANN models. Zhu et al. [14] developed a DNN-based MIR model consisting of four convolution layers and one fully connected layer to perform recognition of transitional motions via a soft lower limb exoskeleton robot with four IMUs and a load cell. This model achieved recognition rates as high as 97.64% when only five steady-state movements (LW, SA, SD, RA, RD) were considered, and the average recognition delay was 23.97% for eight locomotion mode transitions (LW→SA, LW→SD, SA→LW, SD→LW, LW→RA, LW→RD, RA→LW and RD→LW). Narayan et al. [22] used seven IMUs and designed a CNN-based classifier to perform hierarchical classification for lower limb locomotion modes of different specificities and recognition of mode transitions. The model used 1280 milliseconds (ms) of time history data and achieved 94.34% accuracy in the classification of 16 different lower limb locomotion modes. On the other hand, Son and Kang [23] aimed at achieving high-performance human locomotion activity recognition across a variety of terrains and developed a CNN model architecture for the identification of five locomotion activities (LW, SA, SD, RA, RD) leveraging EMG signals measured during the operation of a wearable lower limb robot equipped onto a user and left/right hip angles/velocities and postural signals obtained from the wearable robot. The study used prospective data collected from 500 healthy adult participants and evaluated the impact of the selective use of sensor modalities on the performance of locomotion activity recognition. The results revealed that the use of only biomechanical signals, such as left/right hip angles and velocities and postural signals, resulted in superior performance (96.27% accuracy and 96.17% macro F1 score) and cost effectiveness compared with the use of sEMG signals in addition to biomechanical signals. Several review studies [1], [11] have described challenges related to lower limb MIR and the cooperative control of exoskeleton robots.

B. HUMAN ACTIVITY RECOGNITION

Human activity recognition (HAR) refers to the automatic recognition and classification of many different human actions on the basis of sensor data. HAR shares many similarities with human locomotion recognition (HLR) but is differentiated in terms of its scope and applications. Whereas HLR focuses on locomotion activities such as walking and stair ascent, HAR covers a wider range of activities, such as running and exercising, and locomotion in daily life. Although both share a common feature in that they aim to analyze human motions using wearable sensor data, HAR focuses on the recognition of much more complex and diverse activities. In recent research, HAR has made remarkable achievements in the recognition of complex activities with hybrid models that combine heterogeneous architectures, including CNNs, recurrent neural networks (RNNs), and their variants.

C. CNN-RNN-BASED HYBRID HAR MODELS

Ronao and Cho [25] designed a deep CNN architecture whose components are three convolutional layers, a fully connected layer, and a Softmax layer; the network performs automatic extraction of temporal–local dependency, scale invariance, and hierarchical characteristics of complex human activities. Moreover, Ordóñez and Roggen [26] proposed a deep framework based on CNN and LSTM recurrent units, in which four convolutional layers and two LSTM layers are combined for multimodal activity recognition. The framework was evaluated on two datasets, i.e., OPPORTUNITY and Skoda, and demonstrated recognition accuracies of 91.5% and 95.8%, respectively, which outperformed several existing machine learning methods developed on the basis of hand-crafted features. Xia et al. [27] designed an LSTM−CNN architecture that combines two LSTM layers, two convolution layers,

global average pooling (GAP), and batch normalization to reduce the number of model parameters and speed up network convergence. Notably, the model also achieved an F1 score of 92.71% on the Opportunity dataset, even though it uses fewer model parameters than those required by the deep CNN−LSTM framework [26].

Gupta [28] proposed a hybrid classifier consisting of two CNNs and two stacked gated recurrent unit (GRU) layers that can extract local or temporal features from sensor data of different modalities, such as gyroscope and accelerometer data. This hybrid model achieved recognition accuracies ranging from 90.44% to 96.54% on datasets collected from different mobile devices and smartwatches. On the basis of these achievements in past HAR studies, CNN−GRU architectures with three heads (or branches) [29], [30], [31], which are capable of capturing temporal–local dependencies from sensor data of different modalities, have been proposed. Notably, the models developed by [30], [31] have similar architectures that use three identical convolutional kernel sizes (3, 5, 7) but differ in their extraction of temporal/local features.

D. RESNET-BASED HYBRID HAR MODELS

Zhongkai et al. [32] performed a comparative study on 14 CNN backbone architectures, developing advanced models by embedding three different RNN-based submodules and three different attention mechanisms. The key insights from their comparative study were as follows: (1) In selecting a backbone architecture, many learning parameters do not necessarily lead to improved performance, and the number of learning parameters and layers may have a relatively low impact on the model performance; (2) With respect to embedding submodules, the use of attention modules and, in particular, the squeeze-and-excitation (SE) block, which has a channel-attention mechanism, may be more advantageous for improving model performance than embedding RNN-based submodules; and (3) when lightweight backbone architectures such as MobileNet and the mobile neural architecture search network (MnasNet) are used, the decision of whether to embed an SE block or otherwise should be carefully considered.

For the recognition of transitional activities, which tend to be infrequent and have short durations, Mekruksavanich et al. [33] proposed the SEResNet-BiGRU model, which is composed of eight residual blocks, including SE blocks, and combined it with a bidirectional GRU (BiGRU) for residual mapping of each block. In experiments conducted on the Human Activities and Postural Transitions (HAPT) and MobiAct v2.0 datasets, the proposed hybrid model outperformed three other SOTA hybrid models and exhibited high accuracies of 98.03% and 98.92%, respectively. Mekruksavanich et al. [34] also developed a lightweight deep neural network similar to the aforementioned model but composed of five residual blocks, with a focus on addressing complex HAR problems, for the recognition of specific human behaviors (e.g., eating, drinking, or smoking) in a variety of contexts. The developed network was evaluated on three public datasets: the WISDM-HAR benchmark (WISDM-HARB), UT-Smoke, and UT-Complex. The proposed model outperformed CNN, LSTM, and CNN–LSTM hybrid models in complex HAR tasks and achieved accuracies of 94.91%, 98.75%, and 97.73%, respectively, when approximately 90,000 parameters were used.

To improve feature representation ability for sensor-based HAR tasks, Gao et al. [35] proposed a dual attention network that blends channel and temporal attention mechanisms in a ResNet. In this network, channel and temporal attention are used to determine which channel to focus on among various sensor modalities on different parts of the human body and which sequence to focus on for a target activity. The researchers highlighted that unlike temporal attention mechanisms, in which all sensor modalities are processed without determining which modality is more important, the dual attention mechanism allows automatic learning of the priorities of different sensors, which contributes to superior performance in feature fusion. However, the model had to use 950,000 to 3,510,000 parameters for five datasets (i.e., WISDM, UniMiB SHAR, PAMAP2, OPPORTUNITY, and weakly labeled HAR).

Han et al. [36] indicated that in existing channel-attention modules (i.e., SE blocks), GAP is used to squeeze channel information to minimize computational overhead; however, this process may neglect temporal-aware and modality-aware (TAMA) information, which is actually the information crucial for activity recognition. To address such limitations, their study proposed the use of TAMA attention, in which GAP is factorized into two parallel activity feature embedding processes, emphasizing the relative importance of temporal steps on the basis of activity and sensor modalities at different body positions. Compared with the results obtained by [35], the use of TAMA attention yielded slightly lower model performance but required fewer model parameters, ranging between 320,000 and 880,000 for the four datasets, thus exhibiting superior efficiency.

In contexts similar to these studies, Yang et al. [37] reported that human activities are complex and diverse, encompassing various types of features, and that there are variations in the features emphasized for each activity. To overcome this challenge, researchers have proposed a framework called the multifeature combining attention neural network (MFCANN), in which different convolutional components are stacked in parallel to extract a broader range of features from human activity data. To improve its ability to distinguish between different activities, the network is composed of an intramodule attention block (Intra-MAB) and an intermodule attention block (Inter-MAB), which extract local fine-grained features within feature maps and global distinguishing features across the feature maps, respectively. The performance of this model was compared with those of nine other models, including ResNet, InceptionTime, and Transformer, on the UCI-HAR, the University of Southern

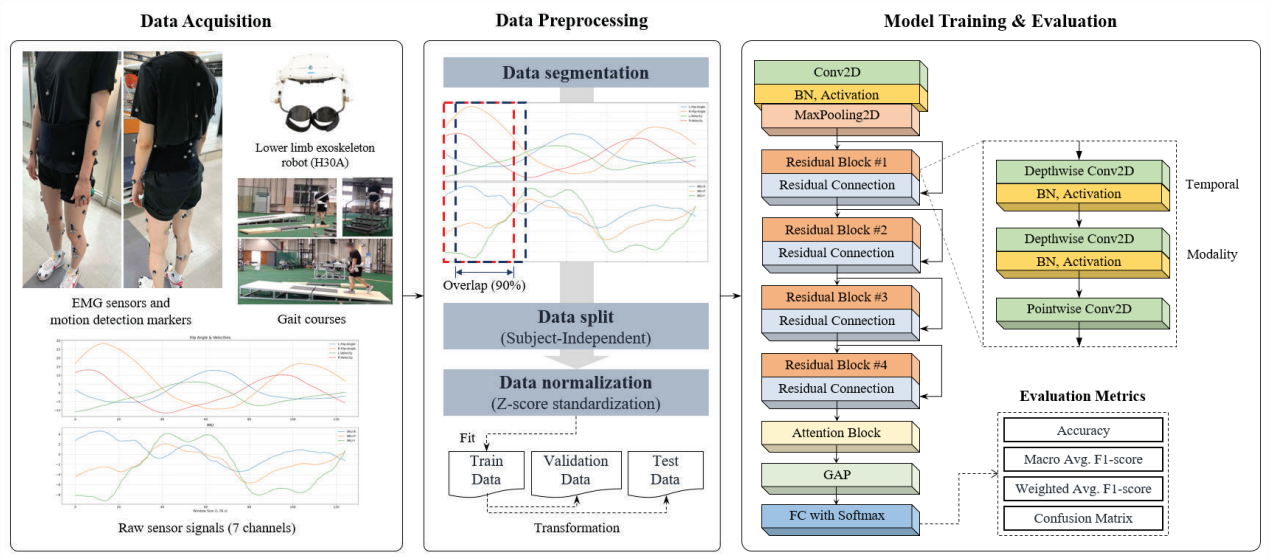

FIGURE 1. Overall workflow of the proposed method.

California: Human Activity Dataset (USC-HAD), and the real-world dataset; the F1 scores were 98.13%, 93.24%, and 99.3%, respectively. Other inception-network-based hybrid models that have been developed recently include the inception-inspired CNN–GRU [38], CNN−BiGRU [39], and inception–attention with GRU [40] models, which have made notable contributions in their attempts to distinguish a range of complex human activities.

III. METHODS

This section presents the workflow of the proposed method for solving the HLR problem via multivariate time series data from a wearable lower-limb exoskeleton robot. Initially, a description of the Walking Assist Wearable Robot Motion (WAWRM) dataset used in the study is presented. The data preprocessing techniques employed in the proposed methods are subsequently described. Finally, our model structure and its corresponding hyperparameters that result in optimal performance are presented. The overall workflow is presented in Fig. 1.

A. WAWRM DATASET

In this study, the proposed method was validated via the WAWRM dataset [23], which consists of prospective data collected from 500 healthy adults (250 men and 250 women) aged between 19 and 64 years, gathered from September 1 to November 30, 2022. The dataset was collected in a previous study involving human subjects. All participants provided written informed consent prior to participation, and the study was approved by the Institutional Review Board (IRB) at Gyeongsang National University Hospital, South Korea (No. GNUCH 2022-08-007-001).

The dataset includes biomechanical signals such as left and right hip joint angles (inches), angular velocities (rpm), postural (roll, pitch, and yaw), and sEMG signals, recorded while participants performed five types of locomotion behaviors (LW, SA, SD, RA, and RD) across three different terrains: flat ground, stairs, and ramps. These activities were carried out while the participants wore a wearable lower limb exoskeleton robot.

The lower limb exoskeleton robot, Hector H30A (HEXAR Humancare, South Korea) [41], was designed to assist the hip joint's muscle strength during locomotion over various terrains. The robot, weighing approximately 4.3 kilograms kg, includes actuators, control units, sensors, and a battery. It features two brushless DC motors that can each deliver up to 12 Newton meters (N·m) of torque to the hip joints. For sensor integration, the robot is equipped with rotary encoders− embedded in the actuator modules − to measure the angles and velocities of the hip joints, as well as an IMU containing a tri-axial accelerometer and a tri-axial gyroscope for estimating posture. The collected biomechanical signals, sampled at a frequency of 71.42857 Hz, include 7-channel wireless data measured at the lowest level (i.e., default mode) of the three torque support modes provided by the robot for hip joint assistance.

Moreover, the sEMG signals were acquired using an 8-channel wireless EMG system (Delsys Trigno, Delsys, Inc., Boston, MA, USA) [42], which was sampled at 2,000 Hz. The electrodes were attached to four muscles on both lower limbs: the vastus lateralis (VL), tibialis anterior (TA), biceps femoris (BF), and gastrocnemius lateralis (GAL). To capture kinematic motion data, an eight-camera motion capture system (Kestrel 2200, Motion Analysis Corp., Santa Rosa, CA, USA) [43] was employed, which records at a sampling rate of 100 Hz. This system tracks the movements of body parts, including the spine, shoulders, elbows, hands, feet, and ankles.

In light of prior findings suggesting limited performance gains or cost efficiency with the inclusion of sEMG

signals [23], we utilized only the seven biomedical signals acquired from the lower-limb exoskeleton robot.

B. DATA PREPROCESSING

A sliding window technique [44] was employed to segment the sensor signals into fixed-length sequences. To determine an appropriate window size – while maintaining a fixed overlap rate of 90%, – the mean (1.5837 s) and standard deviation (0.1743 s) of the left heel strike (LHS) intervals, obtained from 500 healthy adult participants, were used as reference values [23], where s denotes seconds. On this basis, the raw signals are divided into overlapping segments, denoted as $D = \{D_1, D_2, \ldots, D_n\}$, where n is the sequence index and each $D_j \in \mathbb{R}^{T\times S}$ represents a two-dimensional matrix with window lengths of T and S sensor modalities.

Following segmentation, the data were split into training, validation, and test sets. Although random splitting is commonly used, it can lead to information leakage [45], for example, when sequences from the same subject are present in multiple subsets, potentially inflating performance estimates. To mitigate this issue, a subject-independent splitting strategy was adopted, ensuring that sequences from a given participant appeared exclusively in one of the three subsets. In line with a previous study [23], the data were divided into training (400 subjects), validation (50 subjects), and test (50 subjects) sets, enabling evaluation across five types of lower limb locomotion behaviors. Next, to ensure that the model evaluation remained unbiased, Z score standardization [36] was applied by training data statistics, thereby minimizing the influence of anomalous samples. Notably, although signal filtering techniques such as Butterworth filtering can reduce noise, they were avoided in this study to prevent any artificial enhancement of classification performance due to loss of signal variability. Finally, to investigate how model performance varies with window length, the standard deviation used in determining window size was scaled by a factor of $\sigma \in [0, 3]$, and corresponding changes in recognition accuracy were observed.

C. MODEL ARCHITECTURE

To efficiently process multivariate biomechanical signals, the proposed network adopts a lightweight residual architecture that integrates depthwise separable convolutions and an attention module.

1) CONVOLUTIONAL BLOCK

The proposed network employs an initial convolutional block to extract low-level features from segmented input data. As illustrated in Fig. 1, this block is composed of four sequential layers: a 2D convolutional layer (Conv2D), batch normalization (BN), an activation function, and a 2D max pooling layer.

The input tensor is denoted as $\boldsymbol{X} = [x_1, x_2, \ldots, x_C] \in \mathbb{R}^{T\times S\times C}$, where T, S, and C represent the window length (i.e., timesteps), the number of sensor modalities, and the number of channels (initially, $C = 1$), respectively. This tensor is transformed into an initial feature map $\boldsymbol{X}' \in \mathbb{R}^{T'\times S'\times C'}$ through a convolution operation. To enhance training stability by mitigating internal covariate shifts, a BN is applied to the convolution outputs. The scaled exponential linear unit (SeLU) activation function [46] is subsequently employed to improve model expressiveness and address the dying rectified linear unit (ReLU) problem. The values of α and λ were set to 1.6733 and 1.0507, respectively. Although the SeLU exhibits self-normalizing properties, its combination with a BN is intended to foster robust activation dynamics and potentially enhance model performance. The SeLU function, defined as follows, is employed in this study.

$$\mathrm{SeLU}(x) = \begin{cases} \lambda x, & \text{if } x > 0 \\ \lambda\alpha\,(e^x - 1), & \text{if } x \le 0 \end{cases} \tag{1}$$

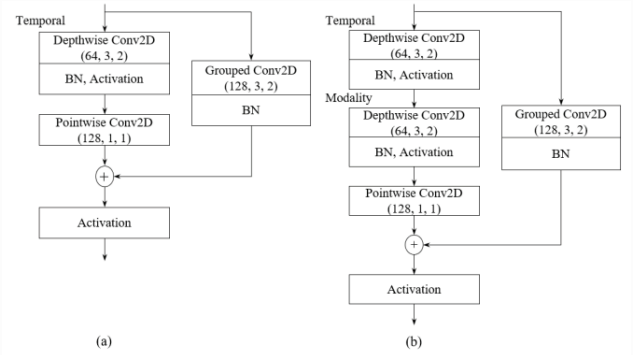

FIGURE 2. Two types of residual blocks with grouped residual connections. (a) Residual block with temporal awareness only. (b) Residual block with both temporal and modality awareness. Each layer is labeled with the number of filters, kernel size, and strides.

2) RESIDUAL BLOCKS

The proposed residual blocks adopt an efficient architecture design to enhance the representation of the feature map extracted from the initial convolutional block. Each block consists of a depthwise convolution [47] for the sequential fusion of temporal and modality-specific features, followed by a pointwise convolution [47] to model interchannel dependencies.

To evaluate the impact of incorporating modality-aware learning in addition to temporal-aware learning within residual blocks, two variants were designed, as illustrated in Fig. 2. Fig. 2(a) depicts a residual block that extracts only temporal features. In contrast, Fig. 2(b) shows a residual block that captures temporal and modality features by employing asymmetric convolutions [48] inspired by Inception-V2. This structure facilitates sequential feature fusion across both the temporal and modality axes. Compared with the former, the latter introduces additional representational capacity along the modality axis, enabling more effective learning of complex locomotion behaviors. In standard residual blocks, a residual connection is used to learn the difference between the input and output feature maps rather than simply passing

the output from the previous block through the convolutional layers. In this study, we adopt the standard residual connection proposed by He et al. [49], as defined in (2), and extend it when the number of output feature map channels increases.

$$y = x + F(x) \tag{2}$$

Specifically, the output of a grouped convolution [50] is added to the temporal or temporal-modality layer output to implement residual mapping. This grouping strategy divides the input channels into multiple subsets, enabling independent convolution operations. Within each group, highly correlated channels are selected to increase learning efficiency.

3) ATTENTION BLOCK

The attention mechanism is widely employed in many fields of deep learning and is useful for capturing the relative importance between pieces of input information and quantifying the importance of each piece of input information on the basis of its characteristics. Notably, the SE block [51] offers advantages in building lightweight units, thus minimizing the computational overhead while improving model performance [32]. The SE block consists of two main steps. The first step, squeezing, aims to embed the global information of each channel into a channel descriptor in the form of a single numeric value (scalar). This process can be represented as a GAP operation that pools the global information across all channels. In (3), z_c denotes the global information of the c-th channel.

$$z_c = \frac{1}{T \times S} \sum_{i=1}^{T} \sum_{j=1}^{S} x_c(i, j) \tag{3}$$

The second step of the SE block, excitation, focuses on learning channel-specific dependence fulfilling the following two criteria on the basis of the information pooled from the squeeze step: (a) learning interchannel, nonlinear interactions should be possible; and (b) learning nonexclusive relationships, where multiple channels can be emphasized at the same time, should be possible. To this end, a reduction rate r is applied via the simple gating mechanism represented in (4), and then rescaling of the final output is performed according to (5):

$$g = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 z)) \tag{4}$$

$$\boldsymbol{Y} = \boldsymbol{X} \times g \tag{5}$$

where $g \in \mathbb{R}^{1\times1\times C}$ is the attention vector calculated via two fully connected layers and where $\boldsymbol{W}_1 \in \mathbb{R}^{C/r\times C}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{C\times C/r}$ represent their respective weight matrices. Moreover, σ denotes the sigmoid function, and δ indicates the ReLU activation function. In the proposed network, the SE block is connected to the end of the last residual block to emphasize important parts of the temporal and modality features. The GAP layer subsequently reduces the spatial dimension of the final feature map while also reducing the possibility of overfitting [52].

4) CLASSIFICATION

This study applied a fully connected layer with a Softmax activation function to compute the class scores for each locomotion behavior. The locomotion behavior was then classified into the class corresponding to the highest predicted probability. During the model training process, the cross-entropy loss function was used to minimize the discrepancy between the predicted and actual labels.

$$\text{loss} = -\sum_{i=1}^{K} y_i \log \hat{y}_i \tag{6}$$

where $\hat{y}_i$ represents the predicted probability of locomotion behavior, y_i is the true label for locomotion behavior, and K is the total number of locomotion behaviors.

5) MODEL AND TRAINING CONFIGURATION

In the proposed network, hyperparameters were carefully selected to balance model complexity and performance. A comprehensive ablation study was conducted to evaluate the contribution of each architectural and training component, including the number of residual blocks, the presence of residual connections, the reduction ratio in channel attention, sensor modalities, and activation functions. The hyperparameter settings are summarized in Table 1.

TABLE 1. The proposed network hyperparameters.

Stage	Parameter	Value
Convolutional block	Conv2D	(64, 5×1, 2×1)
	Max-Pooling2D	(2×1, 2×1)
Residual block1	Depthwise Conv2D (Temporal)	(64, 3, 1)
	Depthwise Conv2D (Modality)	(64, 3, 1)
	Pointwise Conv2D	(64, 1, 1)
Residual block2	Depthwise Conv2D (Temporal)	(64, 3, 1)
	Depthwise Conv2D (Modality)	(64, 3, 1)
	Pointwise Conv2D	(64, 1, 1)
Residual block3	Depthwise Conv2D (Temporal)	(64, 3, 2)
	Depthwise Conv2D (Modality)	(64, 3, 2)
	Pointwise Conv2D	(128, 1, 1)
Residual block4	Depthwise Conv2D (Temporal)	(128, 3, 1)
	Depthwise Conv2D (Modality)	(128, 3, 1)
	Pointwise Conv2D	(128, 1, 1)
SE block	Reduction rate (r)	4–32

Convolutional layer parameters are specified as (number of filters, kernel size, strides), and max-pooling layer as (pool size, strides).

The initial learning rate was set to $1e$-3, and training was performed for 100 epochs. During training, optimization was performed via the Adam optimizer with a fixed minibatch size. The optimal network weights were selected on the basis of the epoch at which the lowest validation loss was observed, whereas the learning rate was reduced by a factor of 0.9 if the validation loss did not improve for 10 consecutive epochs. Although a fixed batch size was used in the main training configuration, additional experiments with five different batch sizes (i.e., 16, 32, 64, 128, and 256) were conducted to assess the impact of this parameter on model performance.

6) EVALUATION METRICS

This study used accuracy to evaluate overall classification performance and the F1 score to fairly assess the performance on locomotion behavior samples. The precision (P) is defined as $TP/TP+FP$, and the recall (R) is defined as $TP/TP+FN$, where TP, FP, FN, and TN denote true positive, false positive, false negative, and true negative, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

For multiclass classification problems, macro- and weighted averaging methods are commonly used to provide a balanced view of model performance across all classes. The macro averaged F1 score is given by (8), where P_i and R_i represent the precision and recall, respectively, for the i-th class:

$$\text{MacroF1} - \text{score} = \frac{1}{K} \sum_{i=1}^{K} \frac{2 \cdot P_i \cdot R_i}{P_i + R_i} \tag{8}$$

By contrast, the weighted average F1 score accounts for class imbalance by weighting each class's F1 score by its support. This is calculated as

$$\text{WeightedF1} - \text{score} = \sum_{i=1}^{K} \frac{N_i}{N} \cdot \frac{2 \cdot P_i \cdot R_i}{P_i + R_i} \tag{9}$$

where N_i is the number of true instances in class i and N is the total number of instances across all classes.

Confusion matrices were also employed to qualitatively interpret the classification performance for each class. Furthermore, model efficiency was evaluated by measuring the number of trainable parameters, as well as the total training and inference times, all expressed in seconds, at the optimal batch size.

IV. EXPERIMENTS

This section presents the experimental setup and compares the proposed model with four competitive methods on the WAWRM dataset. Additionally, a comprehensive ablation study is conducted to analyze the impact of different architectural configurations within the proposed model. To further evaluate the generalizability of the proposed approach in the HAR domain, its performance was assessed on four publicly available benchmark datasets: UCI-HAR [53], HAPT [54], PAMAP2 [55], and WISDM [56].

A. EXPERIMENTAL SETUP

The experiments were conducted on a workstation equipped with an AMD Ryzen 9 3900X 12-core CPU (3.79 GHz), an NVIDIA GeForce RTX 2080 Ti GPU, and 128 GB of RAM. The system operated on Windows 10, and all the implementations were developed using Python 3.8. The deep learning models were built and trained using TensorFlow 2.5, with CUDA 11.2 and cuDNN 8.1.0 for GPU acceleration.

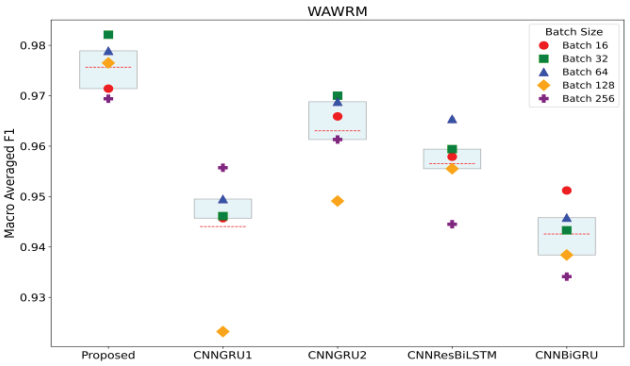

FIGURE 3. Comparison of macro averaged F1-scores for different models on the WAWRM dataset.

TABLE 2. Performance comparison at optimal batch size on the WAWRM dataset.

Model	CNN GRU1 (256)	CNN GRU2 (32)	CNN ResBi LSTM (64)	CNN BiGRU (16)	Proposed (32)
Accuracy	0.9561	0.9702	0.9661	0.952	0.9823
Macro F1	0.9557	0.97	0.9654	0.9512	0.9821
Weighted F1	0.956	0.9701	0.9659	0.9518	0.9823
Trainable parameters	267333	623662	285430	355313	48037
Training time	503.9	3314	2430.7	7735.7	1244.2
Inference time	0.73	2.29	1.7	4.05	0.74

B. COMPARISON METHODS

To assess the effectiveness of the proposed model, its performance was compared with that of four deep learning-based methods that have been extensively utilized in HAR research.

The first group of baseline models consists of CNNGRU1 and CNNGRU2, two hybrid architectures [30], [31] that employ multiscale convolutional layers to extract local temporal features from time series data, followed by GRU layers and feature fusion to enhance representation learning. The second method, CNNResBiLSTM [57], combines a one-dimensional convolutional neural network (1D-CNN), a residual-enhanced bidirectional long short-term memory network, and an attention mechanism to capture spatial and long-range temporal dependencies jointly. The final method, CNNBiGRU [39], is a lightweight architecture that integrates an inception-residual module with bidirectional GRUs to classify activities efficiently via inertial data collected from body-mounted smart devices. These models were selected as baselines because they have proven effective in previous HAR studies using various public datasets. For objective performance comparison, the original structural configurations of the four baseline models were retained, excluding the learning scheduling strategy used for the proposed method.

C. RESULTS OF WAWRM DATASET

As described in Section II, the WAWRM dataset was partitioned using a subject-independent strategy to evaluate

five types of lower limb locomotion behaviors. To create the sequences, the dataset was split using a fixed window length of 1.76 s (125 samples) and a 90% overlap ratio [23]. The dataset consists of 400 subjects (32,951 sequences) for training, 50 (4,120 sequences) for validation, and 50 (4,126 sequences) for testing.

Fig. 3 illustrates the variation in macro F1 scores across different batch sizes for each model on the test set. The macro F1 scores ranged from 0.9232 to 0.9557 for CNNGRU1, 0.9491 to 0.97 for CNNGRU2, 0.9445 to 0.9654 for CNNResBiLSTM, 0.9341 to 0.9512 for CNNBiGRU, and 0.9694 to 0.9821 for the proposed model, indicating that the proposed model consistently achieves higher and more stable performance across batch sizes compared to the baseline models. In terms of overall performance, the average and standard deviation of the macro F1-scores for each model were as follows: CNNGRU1 (0.944 ± 0.0123), CNNGRU2 (0.963 ± 0.0085), CNNResBiLSTM (0.9565 ± 0.0077), CNNBiGRU (0.9426 ± 0.0066), and the proposed model (0.9757 ± 0.0052). Furthermore, the number of trainable parameters in the proposed model (48,037) was reduced by factors ranging from 5.5 to 12.9 compared with those of CNNGRU1 (267,333), CNNGRU2 (623,662), CNNResBiLSTM (285,430), and CNNBiGRU (355,313), as shown in Table 2. This reduction enables faster training and inference times and makes the proposed model more suitable for deployment in resource-constrained environments such as wearable exoskeleton systems while still achieving superior accuracy and stability.

Fig. 4 shows the changes in the validation accuracy and loss over 100 training epochs when the optimal batch size for each model is used. The proposed model outperformed the other four models by achieving the highest validation accuracy and the lowest loss, while also demonstrating more stable performance throughout the training process. These results suggest that the proposed model not only learns more effectively during training but also generalizes better to test data (see Table 2), which is crucial for reliable deployment in real-world HAR applications.

Fig. 5 presents the confusion matrices for the identification of five lower limb locomotion behaviors. Among all the models, the lowest recognition performance was consistently observed for the RD behavior. The misclassification rates from RD to RA were 5.48% for CNNGRU1 and CNNGRU2, 5.62% for CNNResBiLSTM, 4.9% for CNNBiGRU, and 3.6% for the proposed model, which achieved the lowest misclassification rate among them.

D. ABLATION STUDIES

1) RESIDUAL BLOCK

The first ablation study aimed to determine the optimal number of residual blocks. Performance was evaluated under two conditions: (1) when only temporal features were used in each residual block (excluding modality-specific processing, as described in Table 1), and (2) when both temporal and modality features were incorporated.

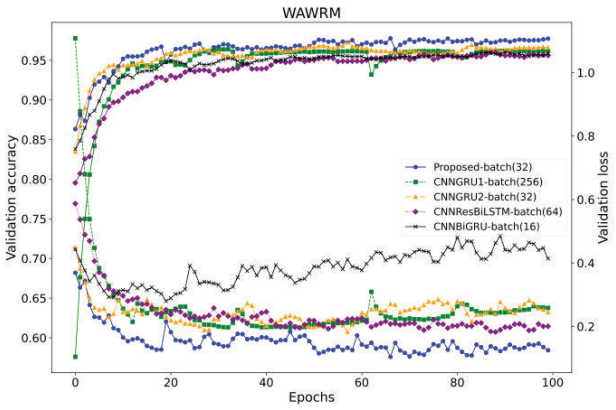

FIGURE 4. Comparison of validation accuracy and loss on the WAWRM dataset.

TABLE 3. Impact of the number of residual blocks on temporal and modality-aware configurations.

Residual block		Accuracy	Macro F1	Weighted F1
Temporal	1	0.7588	0.759	0.7593
	2	0.8349	0.835	0.834
	3	0.8817	0.8826	0.8817
	4	0.9406	0.9413	0.9407
Temporal and modality	1	0.8662	0.8641	0.864
	2	0.954	0.9536	0.9539
	3	0.9658	0.9656	0.9658
	4	0.9748	0.9746	0.9748

The results, summarized in Table 3, show that incorporating both feature types led to consistent improvements in all evaluation metrics as the number of residual blocks increased.

Notably, the highest macro F1 score of 0.9746 was achieved when both temporal and modality features were learned within each block, representing a 3.33% improvement over the configuration using only temporal features. This result underscores the complementary nature of temporal and modality features and demonstrates the effectiveness of their joint learning in enhancing recognition performance.

2) RESIDUAL CONNECTION

We investigated the impact of the presence or absence of residual connections and two types of residual connections on the performance of the proposed model.

As presented in Table 4, the model without residual connections achieved a macro F1 score of 0.9757. Applying standard residual connections yielded a slight improvement of 0.08%, reaching 0.9765, and employing residual connections with grouped convolution led to a substantial improvement of 0.64%, achieving a macro averaged F1 score of 0.9821. The superior performance of residual connections with grouped convolution is likely due to the depthwise and pointwise convolutional structure of the proposed block, which facilitates channelwise feature learning and reduces redundancy. This enables more expressive residual pathways compared to standard residual connections, thereby

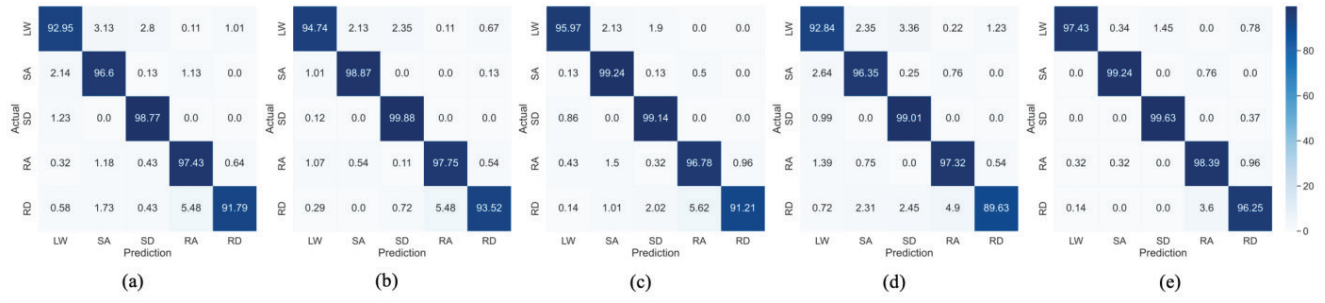

FIGURE 5. Comparison of model confusion matrices. (a) CNNGRU1 (256). (b) CNNGRU2 (32). (c) CNNResBiLSTM (64). (d) CNNBiGRU (16). (e) Proposed model (32). In each model, the value in parentheses represents the optimal batch size.

TABLE 4. Impact of the residual connections.

Residual connection	Accuracy	Macro F1	Weighted F1
None	0.976	0.9757	0.976
Standard convolution	0.977	0.9765	0.9769
Grouped convolution	0.9823	0.9821	0.9823

TABLE 5. Impact of the number of residual blocks with adjustments in the reduction rate for the SE block.

Residual block	Reduction rate	Accuracy	Macro F1	Weighted F1
1	4	0.9006	0.8999	0.9002
2		0.9646	0.9643	0.9646
3		0.9619	0.9614	0.9618
4		0.9823	0.9821	0.9823
1	8	0.9055	0.9059	0.9055
2		0.9527	0.9528	0.9528
3		0.975	0.9748	0.975
4		0.976	0.9757	0.976
1	16	0.8999	0.8993	0.8988
2		0.9338	0.9346	0.9341
3		0.9699	0.9695	0.9699
4		0.9714	0.9713	0.9713
1	32	0.882	0.8821	0.8825
2		0.9636	0.9634	0.9636
3		0.967	0.9669	0.967
4		0.9651	0.9647	0.965

improving the model's representation capability and final performance.

3) ATTENTION BLOCK

In the SE block, the reduction rate r is used to control the parameter overhead. In this ablation study, we investigated its potential impact on model performance by varying $r \in (4, 8, 16, 32)$.

Table 5 shows the changes in the performance of the proposed model as the number of residual blocks and the reduction rate were adjusted. The values of the macro F1 score ranged from 0.8999 to 0.9821 at a reduction rate of 4; from 0.9059 to 0.9757 at a reduction rate of 8; from 0.8993 to 0.9713 at a reduction rate of 16; and from 0.8821 to 0.9669 at a reduction rate of 32. A marginal performance decrease was observed at reduction rates of 4 and 32 as the number of residual blocks increased to 3 and 4, respectively, suggesting that excessively shallow or overly compressed channel representations may hinder learning in deeper architectures. Nevertheless, for most values of the reduction rate, the model performance improved with increasing residual blocks. In addition, the performance improved with decreasing reduction rate, and the best performance was achieved at a reduction rate of 4.

TABLE 6. Impact of five activation functions.

Activation Function	Accuracy	Macro F1	Weighted F1
ReLU	0.9484	0.9481	0.9481
Leaky ReLU (α=0.3)	0.9789	0.9786	0.9789
ELU (α=1)	0.9666	0.9666	0.9665
Swish	0.9719	0.9716	0.9718
SeLU	0.9823	0.9821	0.9823

TABLE 7. Impact of the sensor modality.

Sensor modality	Accuracy	Macro F1	Weighted F1
Encoder	0.9181	0.9177	0.918
Postural	0.9542	0.9538	0.9541
All	0.9823	0.9821	0.9823

4) ACTIVATION FUNCTION

An ablation study was conducted to evaluate the impact of five activation functions: ReLU, leaky ReLU, ELU, Swish (also known as SiLU), and SeLU. Among them, SeLU achieved the highest macro F1 score of 0.9821, followed by leaky ReLU (0.9786), Swish (0.9716), ELU (0.9666), and ReLU (0.9481), as summarized in Table 6.

5) SENSOR MODALITY

The impact of sensor modality selection was evaluated, and the results are presented in Table 7.

When only the left/right hip angles and velocities of the exoskeleton robot wearers were selected, the model achieved a macro F1 score of 0.9177. By contrast, using only postural signals results in a macro F1 score of 0.9538. When all the signal types were combined, the model achieved a macro F1 score of 0.9821, representing an improvement of approximately 6.4% compared with using only the left/right hip angles and velocities.

Compared with the previous results (accuracy of 0.9627, macro F1 score of 0.9617, and total of 840,645 parameters)

[23], the proposed model demonstrated improved performance, with increases of 1.96% in accuracy and 2.04% in the macro F1 score. Additionally, the proposed model is more lightweight, with a reduction in parameters by a factor of 17. Although the improvement in accuracy and macro F1 score may seem moderate, it is meaningful in real-world applications of locomotion behavior recognition, where higher recognition performance can lead to more reliable and responsive assistance. In addition, the proposed model requires significantly fewer parameters, making it more suitable for deployment in wearable systems. These results suggest that, for locomotion behavior recognition, postural signals serve as the most important sensor modality, whereas left/right hip angles and velocities contribute to improved model performance.

6) WINDOW SIZE

Determining the optimal sliding window length is crucial for effectively capturing the patterns in multichannel time series data and enhancing the model learning efficiency. To assess accuracy in recognizing five locomotion behaviors, the reference values for the window size were based on the mean (1.5837 s) and standard deviation (0.1743 s) of LHS measured from 500 healthy adult participants. The standard deviation was scaled by a factor of $\sigma \in [0, 3]$ to examine its impact on model performance.

The results are presented in Table 8. With adjustments to the window size, the accuracy ranged from 0.9695 to 0.9823, and the macro F1 score varied from 0.9695 to 0.9821. Notably, when the scaling factor was set to 1, the model achieved its highest performance.

TABLE 8. Impact of window size.

Window size	Accuracy	Macro F1	Weighted F1
1.58 s (112 samples)	0.9723	0.9721	0.9723
1.67 s (119 samples)	0.9712	0.9707	0.9712
1.76 s (125 samples)	0.9823	0.9821	0.9823
1.85 s (132 samples)	0.9796	0.9796	0.9795
1.93 s (137 samples)	0.9695	0.9695	0.9694
2.02 s (144 samples)	0.9738	0.9739	0.9737
2.11 s (150 samples)	0.9801	0.98	0.9801

7) BATCH SIZE

To assess the impact of batch size, we conducted evaluations with batch sizes ranging from 16 to 256. The results are summarized in Table 9. As the batch size was adjusted, the accuracy ranged from 0.9697 to 0.9823, whereas the macro F1 score varied from 0.9694 to 0.9821. The ablation study revealed that a batch size of 32 yielded the best performance for the proposed model architecture.

E. PERFORMANCE COMPARISON ON FOUR HAR DATASETS

To verify the applicability and generalizability of the proposed model within the HAR domain, we evaluated its performance against four competitive methods using four widely used benchmark datasets: UCI-HAR, HAPT, PAMAP2, and WISDM.

TABLE 9. Impact of batch size.

Window size	Accuracy	Macro F1	Weighted F1
16	0.9714	0.9714	0.9714
32	0.9823	0.9821	0.9823
64	0.9792	0.9789	0.9791
128	0.9767	0.9765	0.9767
256	0.9697	0.9694	0.9697

TABLE 10. Detailed descriptions of the four HAR datasets.

Dataset	UCI-HAR	HAPT	PAMAP2	WISDM
Subjects	30	30	9	36
Activities[a]	6	12	18 (11)	6
Sensors[b]	A and G		A, G, M, and HR	A
Sampling frequency (Hz)	50		IMU (100) and HR (~9)	20
Channels[c]	9	6	54 (27)	3
Missing	–		Linear interpolation	Remove
Downsampling (Hz)	–		33.3	–
Segment size	128		171	128
Overlap	64		85	64
Normalization	*Z* score standardization			
Training	21 subjects		7 subjects	Random split (70%)
Test	6 subjects (2, 9, 10, 13, 18, and 24)		1 subject (106)	Random split (30%)
Validation	3 subjects (4, 12, and 20)		1 subject (105)	30% of the training

[a]The value in parentheses indicates the number of activities (excluding rope jumping) used in the experiments.
[b]A, G, M, and HR denote accelerometer, gyroscope, magnetometer, and heart rate, respectively.
[c]The value in parentheses indicates the number of sensor modalities. In the PAMAP2 dataset, all temperature, HR, orientation, and 3D-acceleration (±6 g) data were excluded. In particular, the accelerometer sometimes becomes saturated during high-impact movements (e.g., running) that exceed ±6 g [55].

Table 10 provides detailed information about these datasets. The first five rows summarize key characteristics, including the number of subjects, activity classes, sensor types, sampling rates, and number of channels. The sixth and seventh rows indicate the presence of missing values and whether downsampling was applied. The following two rows describe the sliding window segmentation criteria used in the experiments, and the tenth row outlines the data normalization method. Finally, the last three rows detail how each dataset was split into training, validation, and test sets. For UCI-HAR and PAMAP2, we followed the subject-independent partitioning strategy used in iSPLInception [58]. Moreover, for WISDM, we adopted a subject-dependent stratified sampling approach to alleviate class imbalance caused by varying data distributions across individuals.

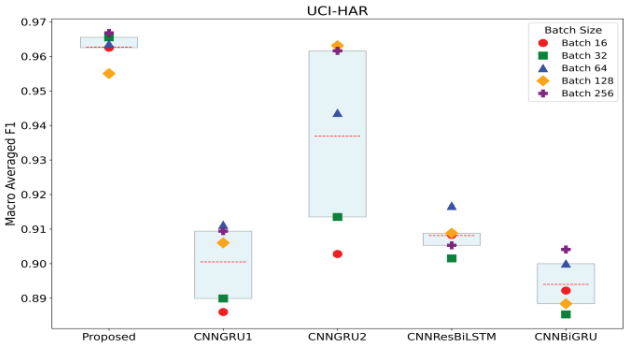

FIGURE 6. Comparison of macro averaged F1-scores for different models on the UCI-HAR dataset.

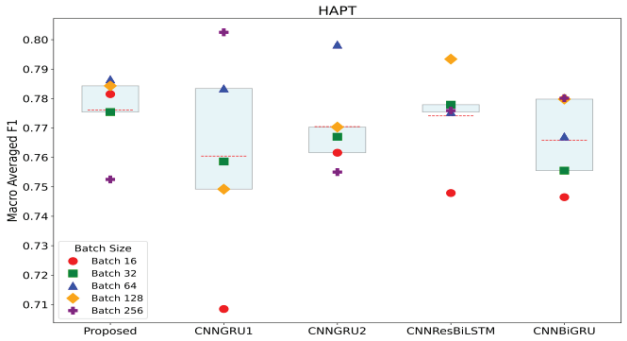

FIGURE 7. Comparison of macro averaged F1-scores for different models on the HAPT dataset.

We also investigated whether the proposed model maintains robust and consistent performance when trained via stochastic gradient descent with warm restarts (SGDR) [59] as an alternative learning strategy. SGDR periodically resets the learning rate to avoid convergence to suboptimal local minima and encourages broader exploration of the optimization space. All the models were trained for up to 100 epochs via the Adam optimizer with an initial learning rate of $1e$-2 and SGDR-specific parameters of $t_{mul} = 1$, $m_{mul} = 0.5$, and $\alpha = 1e$-5, with the learning rate restarts scheduled every 30 epochs.

1) RESULTS ON UCI-HAR DATASET

Table 11 presents the performance of each model on the UCI-HAR dataset, which was evaluated on the basis of the batch size that yielded the best results for each model. Five different batch sizes were tested per model, and the best-performing one was selected. The proposed model achieved the highest performance across all the metrics i.e., accuracy (0.9668), macro F1 score (0.9668), and weighted F1 score (0.9667), while using only 48,166 trainable parameters.

TABLE 11. Performance comparison at the optimal batch size on the UCI-HAR dataset.

Model	CNN GRU1 (64)	CNN GRU2 (128)	CNN ResBi LSTM (64)	CNN BiGRU (256)	Proposed (256)
Accuracy	0.9121	0.9627	0.9172	0.9044	0.9668
Macro F1	0.9112	0.9631	0.9167	0.9041	0.9668
Weighted F1	0.9122	0.9628	0.9172	0.9040	0.9667
Trainable parameters	269318	624205	287098	365202	48166
Training time	500.7	465.9	440.5	309.6	163.7
Inference time	0.44	0.25	0.34	0.46	0.24

Fig. 6 shows the macro F1 score across batch sizes for each model. The macro F1 scores were as follows: CNNGRU1 (0.886–0.9112), CNNGRU2 (0.9028–0.9631), CNNResBiLSTM (0.9015–0.9167), CNNBiGRU (0.8853–0.9041), and the proposed model (0.955–0.9668). Our model presented the highest mean (0.9627) and lowest standard deviation ($\pm$0.0046), indicating strong robustness to batch size changes. This implies lower sensitivity to batch size hyperparameter tuning, enabling more stable training and inference under constrained or variable computational conditions, making it more adaptable to real-world deployment scenarios.

2) RESULTS ON HAPT DATASET

Table 12 summarizes the results on the HAPT dataset, again using the optimal batch size for each model. CNNGRU1 achieved the highest macro F1 score (0.8025), whereas CNNGRU2 achieved the best accuracy (0.9111) and a weighted F1 score (0.9095). Despite having the fewest trainable parameters, the proposed model attained comparable accuracy (0.8939) and a weighted F1 score (0.892) to those of CNNResBiLSTM (0.8939 and 0.8933).

To better understand these results, we further examined the recall performance on the HAPT dataset, focusing on transitional activities. All models demonstrated relatively lower recall on transitions such as standing-to-sitting, standing-to-lying, and lying-to-standing. Compared with CNNGRU2 and CNNResBiLSTM, the proposed model achieved slightly lower recall in the standing-to-sitting task. These results highlight the trade-off between model compactness and detailed motion discrimination.

Fig. 7 illustrates the models' sensitivity to batch size variations. The proposed model (0.776 ± 0.0138), CNNResBiLSTM (0.7741 ± 0.0164), and CNNGRU2 (0.7704 ± 0.0166) displayed relatively stable performance, with the proposed model standing out as the most robust and reliable.

3) RESULTS ON PAMAP2 DATASET

Table 13 presents the performance results on the PAMAP2 dataset when each model's optimal batch size was used. CNNResBiLSTM achieved the highest macro F1 score (0.9182), whereas the proposed model followed closely with 0.9052, despite operating with a smaller batch size. Our model exhibited high efficiency and competitive performance, with 6.2 to 7.4 times fewer parameters than those of CNNResBiLSTM and CNNBiGRU.

We further examined the model's performance on complex daily activities, including ironing, vacuuming, Nordic walking, and running. Compared with CNNResBiLSTM and

TABLE 12. Performance comparison at optimal batch size on the HAPT dataset.

Model	CNN GRU1 (256)	CNN GRU2 (64)	CNN ResBi LSTM (128)	CNN BiGRU (256)	Proposed (64)
Accuracy	0.8923	0.911	0.8939	0.8899	0.8939
Macro F1	0.8025	0.7983	0.7934	0.7801	0.7865
Weighted F1	0.8908	0.9095	0.8933	0.8879	0.892
Trainable parameters	266828	624358	285580	358716	48940
Training time	327.1	809.4	370.5	388.5	254.4
Inference time	0.17	0.51	0.22	0.46	0.13

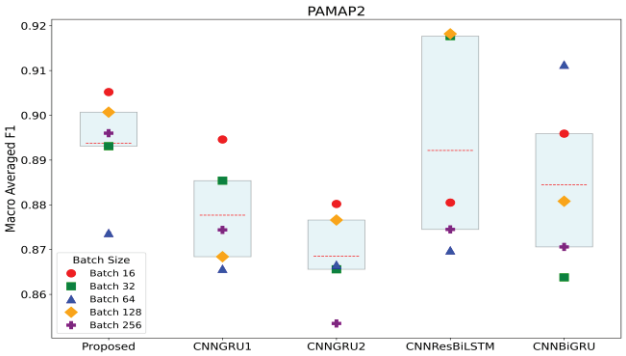

FIGURE 8. Comparison of macro averaged F1-scores for different models on the PAMAP2 dataset.

TABLE 13. Performance comparison at optimal batch size on the PAMAP2 dataset.

Model	CNN GRU1 (16)	CNN GRU2 (16)	CNN ResBi LSTM (128)	CNN BiGRU (64)	Proposed (16)
Accuracy	0.907	0.8896	0.9254	0.9182	0.907
Macro F1	0.8946	0.8802	0.9182	0.9113	0.9052
Weighted F1	0.9055	0.877	0.9231	0.9158	0.9057
Trainable parameters	286923	628576	301579	362151	48811
Training time	1309.7	1593	233.7	549.7	577.7
Inference time	1.03	1.11	0.15	0.23	0.31

CNNBiGRU, the proposed model achieved slightly lower recall in ironing and vacuuming, which may be attributed to the subtle and repetitive nature of upper-body movements in these activities. However, the model maintained comparable performance on more dynamic whole-body activities, such as Nordic walking and running, demonstrating a favorable trade-off between model compactness and recognition capability for high-mobility activities.

These observations are consistent with the overall robustness and efficiency of the proposed model. As depicted in Fig. 8, the proposed model achieved a macro F1 score mean (0.8937 ± 0.0121), with less variation across batch sizes, outperforming all other models in performance and parameter efficiency: CNNResBiLSTM (0.8921 ± 0.0239), CNNBiGRU (0.8845 ± 0.0193), CNNGRU1 (0.8777 ± 0.0121), and CNNGRU2 (0.8685 ± 0.0105).

TABLE 14. Performance comparison at optimal batch size on the WISDM dataset.

Model	CNN GRU1 (256)	CNN GRU2 (128)	CNN ResBi LSTM (128)	CNN BiGRU (128)	Proposed (128)
Accuracy	0.9866	0.9845	0.9897	0.9872	0.9829
Macro F1	0.9803	0.9786	0.9843	0.9808	0.976
Weighted F1	0.9866	0.9845	0.9897	0.9872	0.9829
Trainable parameters	263558	622963	282490	355074	48166
Training time	597.3	903.1	615.8	768.7	273.2
Inference time	0.25	0.56	0.45	0.45	0.15

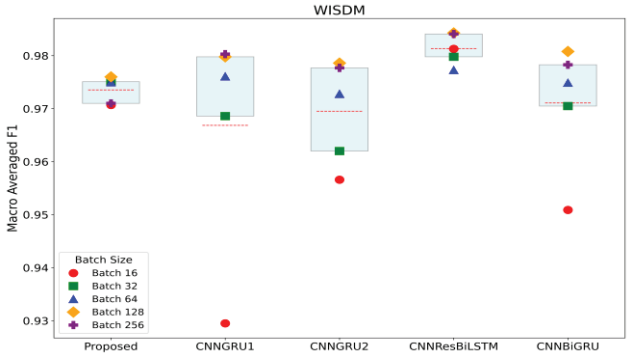

FIGURE 9. Comparison of macro averaged F1-scores for different models on the WISDM dataset.

4) RESULTS ON WISDM DATASET

Table 14 summarizes the best macro F1 scores of all the models on the WISDM dataset.

Although the peak performance of the proposed model was slightly lower than that of the top competitor, it still demonstrated high consistency (0.9736 ± 0.0025), second only to CNNResBiLSTM (0.9814 ± 0.003) and ahead of CNNBiGRU (0.9711 ± 0.0119), as shown in Fig. 9. Importantly, the proposed model required 5.5 to 12.9 times fewer parameters than its competitors, striking a favorable balance between accuracy and computational efficiency. This makes it particularly suitable for deployment in resource-constrained real-world environments.

V. CONCLUSION

In this study, we presented a lightweight ResNet-based model for recognizing diverse locomotion activities via a wearable lower-limb exoskeleton robot. The proposed architecture integrates asymmetrically designed lightweight residual blocks and a channel-attention mechanism to extract temporal and modality-specific features effectively while minimizing the computational cost.

The experimental results on the WAWRM dataset, which was collected from 500 adults via the Hector H30A exoskeleton robot, demonstrated that the model achieved an accuracy of 98.23% and a macro F1 score of 98.21% when only 48,037 trainable parameters were used. Additional evaluations on four public HAR datasets (UCI-HAR, HAPT, PAMAP2,

and WISDM) demonstrated consistent performance and low macro F1 score variance, despite the model's compact size (48,166–48,940 parameters across datasets). These results confirm the model's robustness and generalizability across domains. Overall, the proposed method achieved a favorable balance between model compactness and recognition accuracy, making it suitable for real-time HAR applications in wearable systems with limited computational resources.

In future work, we plan to deploy the proposed lightweight model on a representative embedded system integrated into a wearable exoskeleton robot, evaluating not only its recognition performance under real-time operating conditions but also its inference speed and memory efficiency. As real-time control is a critical component of human-robot interaction, we will also explore incorporating torque profiles—derived from user movements and control signals—into the control strategy to enable more intelligent and adaptive system responses. Furthermore, to extend the model's applicability to more complex HAR tasks involving mixed locomotion and actions, we intend to optimize its architecture through AutoML-based hyperparameter tuning, ensuring robust and scalable performance across diverse real-world scenarios.

REFERENCES

[1] L.-L. Li, G.-Z. Cao, H.-J. Liang, Y.-P. Zhang, and F. Cui, "Human lower limb motion intention recognition for exoskeletons: A review," *IEEE Sensors J.*, vol. 23, no. 24, pp. 30007–30036, Dec. 2023, doi: 10.1109/JSEN.2023.3328615.

[2] A. Rodríguez-Fernández, J. Lobo-Prat, and J. M. Font-Llagunes, "Systematic review on wearable lower-limb exoskeletons for gait training in neuromuscular impairments," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 22, Feb. 2021, doi: 10.1186/s12984-021-00815-5.

[3] G. Chen, P. Qi, Z. Guo, and H. Yu, "Mechanical design and evaluation of a compact portable knee–ankle–foot robot for gait rehabilitation," *Mechanism Mach. Theory*, vol. 103, pp. 51–64, Sep. 2016, doi: 10.1016/j.mechmachtheory.2016.04.012.

[4] L. N. Awad, J. Bae, K. O'Donnell, S. M. M. De Rossi, K. Hendron, L. H. Sloot, P. Kudzia, S. Allen, K. G. Holt, T. D. Ellis, and C. J. Walsh, "A soft robotic exosuit improves walking in patients after stroke," *Sci. Translational Med.*, vol. 9, no. 400, pp. 1–12, Jul. 2017, doi: 10.1126/scitranslmed.aai9084.

[5] S. Au, M. Berniker, and H. Herr, "Powered ankle-foot prosthesis to assist level-ground and stair-descent gaits," *Neural Netw.*, vol. 21, no. 4, pp. 654–666, May 2008, doi: 10.1016/j.neunet.2008.03.006.

[6] F. Sup, H. Atakan Varol, J. Mitchell, T. J. Withrow, and M. Goldfarb, "Preliminary evaluations of a self-contained anthropomorphic transfemoral prosthesis," *IEEE/ASME Trans. Mechatronics*, vol. 14, no. 6, pp. 667–676, Dec. 2009, doi: 10.1109/TMECH.2009.2032688.

[7] V. Monaco, P. Tropea, F. Aprigliano, D. Martelli, A. Parri, M. Cortese, R. Molino-Lova, N. Vitiello, and S. Micera, "An ecologically-controlled exoskeleton can improve balance recovery after slippage," *Sci. Rep.*, vol. 7, no. 1, p. 46721, May 2017, doi: 10.1038/srep46721.

[8] E. Martini, S. Crea, A. Parri, L. Bastiani, U. Faraguna, Z. McKinney, R. Molino-Lova, L. Pratali, and N. Vitiello, "Gait training using a robotic hip exoskeleton improves metabolic gait efficiency in the elderly," *Sci. Rep.*, vol. 9, no. 1, p. 7157, May 2019, doi: 10.1038/s41598-019-43628-2.

[9] S.-H. Lee, J. Kim, B. Lim, H.-J. Lee, and Y.-H. Kim, "Exercise with a wearable hip-assist robot improved physical function and walking efficiency in older adults," *Sci. Rep.*, vol. 13, no. 1, p. 7269, May 2023, doi: 10.1038/s41598-023-32335-8.

[10] J. Baraglia, M. Cakmak, Y. Nagai, R. P. Rao, and M. Asada, "Efficient human–robot collaboration: When should a robot take initiative?" *Int. J. Robot. Res.*, vol. 36, nos. 5–7, pp. 563–579, Feb. 2017, doi: 10.1177/0278364916688253.

[11] G. Masengo, X. Zhang, R. Dong, A. B. Alhassan, K. Hamza, and E. Mudaheranwa, "Lower limb exoskeleton robot and its cooperative control: A review, trends, and challenges for future research," *Frontiers Neurorobotics*, vol. 16, Jan. 2023, Art. no. 913748, doi: 10.3389/fnbot.2022.913748.

[12] L. Zhang, G. Liu, B. Han, Z. Wang, and T. Zhang, "SEMG based human motion intention recognition," *J. Robot.*, vol. 2019, pp. 1–12, Aug. 2019, doi: 10.1155/2019/3679174.

[13] D. Farina, R. Merletti, and R. M. Enoka, "The extraction of neural strategies from the surface EMG," *J. Appl. Physiol.*, vol. 96, no. 4, pp. 1486–1495, Apr. 2004, doi: 10.1152/japplphysiol.01070.2003.

[14] L. Zhu, Z. Wang, Z. Ning, Y. Zhang, Y. Liu, W. Cao, X. Wu, and C. Chen, "A novel motion intention recognition approach for soft exoskeleton via IMU," *Electronics*, vol. 9, no. 12, p. 2176, Dec. 2020, doi: 10.3390/electronics9122176.

[15] K. Li, J. Zhang, L. Wang, M. Zhang, J. Li, and S. Bao, "A review of the key technologies for sEMG-based human–robot interaction systems," *Biomed. Signal Process. Control*, vol. 62, Sep. 2020, Art. no. 102074, doi: 10.1016/j.bspc.2020.102074.

[16] S. Tortora, L. Tonin, C. Chisari, S. Micera, E. Menegatti, and F. Artoni, "Hybrid human-machine interface for gait decoding through Bayesian fusion of EEG and EMG classifiers," *Frontiers Neurorobotics*, vol. 14, Nov. 2020, Art. no. 582728, doi: 10.3389/fnbot.2020.582728.

[17] Z. Ding, C. Yang, Z. Wang, X. Yin, and F. Jiang, "Online adaptive prediction of human motion intention based on sEMG," *Sensors*, vol. 21, no. 8, p. 2882, Apr. 2021, doi: 10.3390/s21082882.

[18] D. Xiong, D. Zhang, X. Zhao, Y. Chu, and Y. Zhao, "Synergy-based neural interface for human gait tracking with deep learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2271–2280, 2021, doi: 10.1109/TNSRE.2021.3123630.

[19] C. Yi, F. Jiang, S. Zhang, H. Guo, C. Yang, Z. Ding, B. Wei, X. Lan, and H. Zhou, "Continuous prediction of lower-limb kinematics from multi-modal biomedical signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2592–2602, May 2022, doi: 10.1109/TCSVT.2021.3071461.

[20] K. Gui, H. Liu, and D. Zhang, "A practical and adaptive method to achieve EMG-based torque estimation for a robotic exoskeleton," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 2, pp. 483–494, Apr. 2019, doi: 10.1109/TMECH.2019.2893055.

[21] B. Xiong, N. Zeng, H. Li, Y. Yang, Y. Li, M. Huang, W. Shi, M. Du, and Y. Zhang, "Intelligent prediction of human lower extremity joint moment: An artificial neural network approach," *IEEE Access*, vol. 7, pp. 29973–29980, 2019, doi: 10.1109/ACCESS.2019.2900591.

[22] A. Narayan, F. A. Reyes, M. Ren, and Y. Haoyong, "Real-time hierarchical classification of time series data for locomotion mode detection," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1749–1760, Apr. 2022, doi: 10.1109/JBHI.2021.3106110.

[23] C.-S. Son and W.-S. Kang, "Multivariate CNN model for human locomotion activity recognition with a wearable exoskeleton robot," *Bioengineering*, vol. 10, no. 9, p. 1082, Sep. 2023, doi: 10.3390/bioengineering10091082.

[24] B. Hu, E. Rouse, and L. Hargrove, "Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities," *Frontiers Robot. AI*, vol. 5, p. 78, Jun. 2018, doi: 10.3389/frobt.2018.00078.

[25] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016, doi: 10.1016/j.eswa.2016.04.032.

[26] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016, doi: 10.3390/s16010115.

[27] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020, doi: 10.1109/ACCESS.2020.2982225.

[28] S. Gupta, "Deep learning based human activity recognition (HAR) using wearable sensor data," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 2, Nov. 2021, Art. no. 100046, doi: 10.1016/j.jjimei.2021.100046.

[29] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, vol. 103, no. 7, pp. 1461–1478, Jul. 2021, doi: 10.1007/s00607-021-00928-8.

[30] L. Lu, C. Zhang, K. Cao, T. Deng, and Q. Yang, "A multichannel CNN-GRU model for human activity recognition," *IEEE Access*, vol. 10, pp. 66797–66810, 2022, doi: 10.1109/ACCESS.2022.3185112.

[31] C. Zhang, K. Cao, L. Lu, and T. Deng, "A multi-scale feature extraction fusion model for human activity recognition," *Sci. Rep.*, vol. 12, no. 1, p. 20620, Nov. 2022, doi: 10.1038/s41598-022-24887-y.
[32] Z. Zhongkai, S. Kobayashi, K. Kondo, T. Hasegawa, and M. Koshino, "A comparative study: Toward an effective convolutional neural network architecture for sensor-based human activity recognition," *IEEE Access*, vol. 10, pp. 20547–20558, 2022, doi: 10.1109/ACCESS.2022.3152530.
[33] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "A hybrid deep residual network for efficient transitional activity recognition based on wearable sensors," *Appl. Sci.*, vol. 12, no. 10, p. 4988, May 2022, doi: 10.3390/app12104988.
[34] S. Mekruksavanich, A. Jitpattanakul, K. Sitthithakerngkiet, P. Youplao, and P. Yupapin, "ResNet-SE: Channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors," *IEEE Access*, vol. 10, pp. 51142–51154, 2022, doi: 10.1109/ACCESS.2022.3174124.
[35] W. Gao, L. Zhang, Q. Teng, J. He, and H. Wu, "DanHAR: Dual attention network for multimodal human activity recognition using wearable sensors," *Appl. Soft Comput.*, vol. 111, Nov. 2021, Art. no. 107728, doi: 10.1016/j.asoc.2021.107728.
[36] C. Han, L. Zhang, Y. Tang, S. Xu, F. Min, H. Wu, and A. Song, "Understanding and improving channel attention for human activity recognition by temporal-aware and modality-aware embedding," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022, doi: 10.1109/TIM.2022.3191653.
[37] Z. Yang, K. Li, and Z. Huang, "MFCANN: A feature diversification framework based on local and global attention for human activity recognition," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108110, doi: 10.1016/j.engappai.2024.108110.
[38] N. Dua, S. N. Singh, V. B. Semwal, and S. K. Challa, "Inception inspired CNN-GRU hybrid network for human activity recognition," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 5369–5403, Feb. 2023, doi: 10.1007/s11042-021-11885-x.
[39] H. A. Imran, Q. Riaz, M. Hussain, H. Tahir, and R. Arshad, "Smart-wearable sensors and CNN-BiGRU model: A powerful combination for human activity recognition," *IEEE Sensors J.*, vol. 24, no. 2, pp. 1963–1974, Jan. 2024, doi: 10.1109/JSEN.2023.3338264.
[40] T. R. Mim, M. Amatullah, S. Afreen, M. A. Yousuf, S. Uddin, S. A. Alyami, K. F. Hasan, and M. A. Moni, "GRU-INC: An inception-attention based approach using GRU for human activity recognition," *Expert Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119419, doi: 10.1016/j.eswa.2022.119419.
[41] HEXAR-Humancare. *Hector H30A*. Accessed: May 2, 2025. [Online]. Available: https://hexarhc.com/?page_id=5465&lang=en
[42] DELSYS. *Trigno Wireless Biofeedback System*. Accessed: May 2, 2025. [Online]. Available: https://delsys.com/support/documentation/#usersguide
[43] Motion Analysis. *Kestrel-2200*. Accessed: May 2, 2025. [Online]. Available: https://www.motionanalysis.com/cameras/kestrel-2200/
[44] A. Dehghani, O. Sarbishei, T. Glatard, and E. Shihab, "A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors," *Sensors*, vol. 19, no. 22, p. 5026, Nov. 2019, doi: 10.3390/s19225026.
[45] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, Sep. 2023, Art. no. 100804, doi: 10.1016/j.patter.2023.100804.
[46] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2017, pp. 972–981.
[47] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[50] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5987–5995, doi: 10.1109/CVPR.2017.634.
[51] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.
[52] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arxiv:13124400*.
[53] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21st Eur. Symp. Artif. Neural Netw. Comput. Intell. Mach. Learn. (ESANN)*, Bruges, Belgium, Apr. 2013, pp. 437–442.
[54] J. L. Reyes-Ortiz, L. Oneto, A. Ghio, A. Samá, D. Anguita, and X. Parra, "Human activity recognition on smartphones with awareness of basic activities and postural transitions," in *Artificial Neural Networks and Machine Learning*. Cham, Switzerland: Springer, 2014, pp. 177–184, doi: 10.1007/978-3-319-11179-7_23.
[55] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Proc. 16th Int. Symp. Wearable Comput.*, Jun. 2012, pp. 108–109, doi: 10.1109/ISWC.2012.13.
[56] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newslett.*, vol. 12, no. 2, pp. 74–82, Mar. 2011, doi: 10.1145/1964897.1964918.
[57] J. Zhang, Y. Liu, and H. Yuan, "Attention-based residual BiLSTM networks for human activity recognition," *IEEE Access*, vol. 11, pp. 94173–94187, 2023, doi: 10.1109/ACCESS.2023.3310269.
[58] M. Ronald, A. Poulose, and D. S. Han, "ISPLInception: An inception-ResNet deep learning architecture for human activity recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021, doi: 10.1109/ACCESS.2021.3078184.
[59] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.

CHANG-SIK SON received the B.Sc., M.Sc., and Ph.D. degrees in computer science from the Catholic University of Daegu, Daegu, South Korea, in 2000, 2002, and 2006, respectively.

From 2007 to 2009, he was a Postdoctoral Fellow with the Department of Electrical Engineering, Yeungnam University, Daegu, South Korea. From 2009 to 2014, he was a Research Fellow with the Department of Medical Informatics, Keimyung University Dongsan Medical Center, Daegu. Since 2014, he has been a Senior Researcher with Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, South Korea. His research interests include knowledge discovery, pattern recognition, artificial intelligence-based human–robot interaction, and biomedical informatics applications.

WON-SEOK KANG received the Ph.D. degree in biomedical science from Kyungpook National University, in 2024, and the B.S. and M.S. degrees in computer science and engineering from Yeungnam University, South Korea, in 1998 and 2000, respectively. He was a Researcher with KAIST, South Korea, from January 2000 to December 2004. From 2019 to 2020, he was a Visiting Researcher with UMASS Lowell, USA. Since August 2005, he has been a Principal Researcher with Daegu Gyeongbuk Institute of Science and Technology, (DGIST), South Korea. His research interests include digital phenotyping, healthcare S/W, biomedical signal/data processing, data mining, machine learning, and simulation & modeling.