symmetry

MDPI

Article

Deep Residual Network with a CBAM Mechanism for the Recognition of Symmetric and Asymmetric Human Activity Using Wearable Sensors

Sakorn Mekruksavanich [1] **and Anuchit Jitpattanakul** [2,3,*]

[1] Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand; sakorn.me@up.ac.th
[2] Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
[3] Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
[*] Correspondence: anuchit.j@sci.kmutnb.ac.th

Abstract: Wearable devices are paramount in health monitoring applications since they provide contextual information to identify and recognize human activities. Although sensor-based human activity recognition (HAR) has been thoroughly examined, prior studies have yet to definitively differentiate between symmetric and asymmetric motions. Determining these movement patterns might provide a more profound understanding of assessing physical activity. The main objective of this research is to investigate the use of wearable motion sensors and deep convolutional neural networks in the analysis of symmetric and asymmetric activities. This study provides a new approach for classifying symmetric and asymmetric motions using a deep residual network incorporating channel and spatial convolutional block attention modules (CBAMs). Two publicly accessible benchmark HAR datasets, which consist of inertial measurements obtained from wrist-worn sensors, are used to assess the model's efficacy. The model we have presented is subjected to thorough examination and demonstrates exceptional accuracy on both datasets. The ablation experiment examination also demonstrates noteworthy contributions from the residual mappings and CBAMs. The significance of recognizing basic movement symmetries in increasing sensor-based activity identification utilizing wearable devices is shown by the enhanced accuracy and F1-score, especially in asymmetric activities. The technique under consideration can provide activity monitoring with enhanced accuracy and detail, offering prospective advantages in diverse domains like customized healthcare, fitness tracking, and rehabilitation progress evaluation.

Keywords: human activity recognition; wearable sensor; symmetric human activity; deep learning; deep residual network

check for updates

Citation: Mekruksavanich, S.; Jitpattanakul, A. Deep Residual Network with a CBAM Mechanism for the Recognition of Symmetric and Asymmetric Human Activity Using Wearable Sensors. *Symmetry* **2024**, *16*, 554. https://doi.org/10.3390/sym16050554

Academic Editor: Shangce Gao

Received: 17 March 2024
Revised: 13 April 2024
Accepted: 22 April 2024
Published: 3 May 2024

1. Introduction

Wearable gadgets enable the ongoing surveillance of people's actions, providing crucial contextual data for well-being and health solutions [1]. Smartwatches and wristbands include inertial measurement units (IMUs) to detect and analyze human movements and recognize human behavior efficiently [2]. Fitness trackers and smartwatches integrated with IMUs have become widely popular in tracking wellness and health. The integration of motion sensors in wearables allows for the continuous identification of human actions, opening up various applications beyond just monitoring activities [3]. For example, human activity recognition (HAR) identifies sedentary activities to avoid lifestyle disorders [4]. Determining improvement over physical treatment may be facilitated by assessing asymmetry in arm motions, which can benefit stroke recovery [5]. Smartwatch-based HAR also shows potential for particular groups, such as older adults, who are prone to falls [6]. The remote assessment of recovery and disability is facilitated by detecting motion anomalies caused by injuries, arthritic conditions, and surgery. The emphasis on monitoring aberrant asymmetry sets HAR apart from typical activity categorization challenges. In addition to developing practical computing on wearable devices, HAR offers individualized biofeedback to prevent the worsening of pre-existing conditions during physical activities. The increasing usefulness of sensor-based HAR algorithms on commercialized smartwatches is evident in their detailed understanding of an individual's mobility, stability, and health issues.

Inertial sensor feeds have been effectively classified for physical activities using deep learning techniques, resulting in high accuracy [7]. Nevertheless, most current HAR techniques fail to differentiate between symmetric activities like walking or jogging and asymmetric actions like playing golf or tossing a ball. Discovering inherent patterns of symmetry or asymmetry in movements may provide profound insights into the underlying meaning linked with activities. Assessing the asymmetry in gait and arm movement could be helpful as an indicator of recovery status in stroke rehabilitation [8].

This research introduces a method that utilizes a deep convolutional neural network (CNN) and attention mechanisms to accurately identify symmetric and asymmetric movements from multi-dimensional inertial information. The main contributions of this study can be stated as follows:

1. This study delves into detecting symmetric and asymmetric human activities utilizing wearable sensors. To achieve this, we employed two well-established benchmark HAR datasets: WISDM-HARB and UTwente. These datasets offer diverse symmetric and asymmetric human actions, providing a robust foundation for our research.
2. The proposed model, CNN-ResBLSTM-CBAM, represents an innovative approach, integrating advanced deep residual networks with attention mechanisms. This design is tailored to effectively learn and capture the nuanced characteristics of symmetry and asymmetry in sensor data.
3. Extensive evaluations demonstrate the efficacy of our method, showcasing impressive accuracy rates of 89.01% and 96.49% on the WISDM-HARB and UTwente datasets, respectively. These evaluations emphasize the model's ability to differentiate between symmetric and asymmetric activities. Notably, our approach surpasses the performance of conventional CNNs and long short-term memory (LSTM) models in this classification task.
4. Furthermore, our study conducts thorough assessments to elucidate the impact of various sensor types on the classification of symmetric and asymmetric human activities. This comprehensive analysis sheds light on the nuances of sensor selection and its implications for accurate activity recognition.

The remaining sections are structured as follows: Section 2 provides an overview of HAR and sensing modalities in wearable inertial sensors and critically evaluates previous deep learning methodologies and their limitations. Section 3 outlines the proposed attention-based deep CNN technique for simulating symmetric and asymmetric movements, highlighting advancements in model components. Section 4 details the experimental setup, dataset, implementation, and benchmarks of the approach against state-of-the-art methods. Section 5 investigates individual model components' contributions and provides qualitative representations of learned properties. Section 6 concludes by reviewing results, addressing limitations, suggesting future research directions, and exploring the implications of practical sensor-driven activity identification systems.

2. Related Works

This section comprehensively summarizes previous studies concerning sensor-based HAR. We thoroughly review existing research on HAR techniques employing wearable sensors to gather movement data. Furthermore, we offer a succinct overview of various deep learning methodologies utilized to improve the performance of HAR systems, such as convolutional and recurrent neural network structures. Our assessment includes an examination of both significant advancements and inherent limitations observed in current deep learning models for activity categorization.

2.1. Sensor-Based HAR

HAR involves using sensor data to identify and classify physical activities carried out by persons. Automated HAR facilitates the implementation of diverse sports, healthcare, military, and entertainment applications, allowing them to adjust and react accordingly to user behavior. Wearable sensors may collect movement data, which is then analyzed by identification systems to detect specific activity, such as walking, jogging, or sitting. The acknowledged action may provide context to customized applications to more effectively cater to unique requirements or preferences in that particular circumstance. In recent years, extensive research has significantly improved sensor-based activity categorization, making it suitable for many developing applications such as remote medical monitoring, tailored fitness advice, interactive gaming, automatic security warnings, and more [9,10]. There are still many opportunities to enhance the development of robust and precise models for recognizing human activities. These models allow the next level of intelligent user-focused services in many sectors.

Previous studies have classified human actions into two categories: simple human activities (SHA) and complex human activities (CHA) [11–13]. According to Shoaib et al. [13], SHAs are repetitive, regular movements such as walking, running, sitting, and standing that can be easily recognized with an accelerometer. On the other hand, CHAs are not as repeated and often include actions associated with the hands, such as eating, smoking, and drinking [11]. Recognizing these actions may require using other sensory modalities, such as a gyroscope. Actions that include stairs have also been categorized as CHAs because of the challenge of accurately identifying them alone via an accelerometer [13]. Alo et al. [14] distinguished short-duration activities, such as walking and jogging, and longer-duration activities, such as cooking or taking medicine, which entail a series of sub-activities. Peng et al. [12] defined SHAs as repetitive movements or individual body positions that, while easier to identify, do not accurately represent typical human activities in everyday life. The complication of real-world activities, including many SHAs, temporal evolution, and high-level interpretations, is best encapsulated by CHAs. Liu et al. [15] defined atomic actions as indivisible unit activities, meaning they cannot be broken down any further. On the other hand, complex activities consist of a series of atomic movements performed in different combinations, either sequentially or simultaneously. Correspondingly, Chen et al. [16] conducted a comparison between single and repetitive SHAs, which can be identified using an accelerometer and CHAs. CHAs involve multiple overlapping activities and often require the utilization of multimodal sensors. Ultimately, Lara and Labrador [11] presented a review that condensed taxonomies of human activities, categorized according to the distinction between primary and complicated motions, as previously defined in existing research.

The concept of symmetry has received limited attention in studies on HAR, as evidenced by [17,18]. Furthermore, it can also be explored concerning indoor activity recognition. For instance, strolling, as an ordinary human action, can be categorized as a symmetrical movement based on features inherent to bipedal locomotion, such as the angles of incline. Similarly, jogging exhibits symmetrical characteristics, as both arms and legs move in synchrony at a consistent rate; essentially, the phase-plane cycles of each leg mirror each other.

2.2. Deep Learning Approaches for HAR

Despite progress, identifying human activities still needs to be completed. Feature extraction plays a crucial role in gathering relevant information and distinguishing between actions, thereby impacting accuracy. Methods for feature extraction can be broadly categorized into two approaches: human feature engineering and automated feature learning using deep learning techniques. The manual engineering of domain-specific features is a laborious and complex process that may not be readily applicable across various activities. On the other hand, deep learning enables the automated extraction of hierarchical features, eliminating the need for extensive domain expertise [19]. In end-to-end HAR models, deep neural networks such as recurrent neural networks (RNNs) and CNNs have been employed to automatically learn features from time series data.

Cutting-edge hierarchical autoregressive HAR models have emphasized crafting sophisticated and deep architectures that extend beyond simple feature learning to improve accuracy. These models commonly leverage CNN-based feature extractors as their primary structural elements. Numerous investigations have introduced tailored backbone architectures, including Inception modules [20], dual-scale residual networks [21], ResNet ensembles [22], and iSPLInception [23]. Additionally, certain researchers have employed DenseNet backbones [24].

Various improvements have been integrated into CNN backbones, incorporating LSTM layers [25–28] to capture temporal patterns and self-attention blocks to selectively enhance features according to contextual information. Examples of models employing these enhancements include combinations like BiLSTMs coupled with Inception-CNN branches [27], CNNs paired with BiLSTMs [28], dual-attention CNN architectures [29], DeepConvLSTMs incorporating attention mechanisms [30], and multi-channel CNNs with embedded attention [31]. These recent studies on HAR that use deep learning technology are given in Table 1.

Table 1. Comparison of the recent HAR studies with our study.

Ref.	Year	Method	Dataset	Sensor Types	Sensor Location	No. of Activities	Focus Symmetry and Asymmetry Motions
[21]	2019	Dual-scaled residual network	OPPORTUNITY UniMiB-SHAR	A, G A	Body Pocket	11 17	no
[22]	2020	Ensemble ResNet	UCI-DSA	A, G, M	Body	19	no
[23]	2021	iSPLInception	OPPORTUNITY PAMAP2	A, G A, G, M	Body Head, Chest, Ankle	17 12	no
[24]	2020	DenseNet	UCI-HAR	A, G	Waist	6	no
[25]	2019	InnoHAR	OPPORTUNITY PAMAP2 UCI-HAR	A, G A, G, M A, G	Body Head, Chest, Ankle Waist	17 12 6	no
[26]	2017	Residual BiLSTM	OPPORTUNITY UCI-HAR	A, G A, G	Body Waist	17 6	no
[27]	2020	BiLSTM	mHelath	A, G, M	Ankle, Arm, Chest	12	no
[28]	2021	Multibrance CNN-BiLSTM	WISDM UCI-HAR PAMAP2	A A, G A, G, M	Pocket Waist Head, Chest, Ankle	6 6 12	no
[29]	2021	Dual Attention Network	WISDM UniMiB-SHAR PAMAP2 OPPORTUNITY	A A A, G, M A, G	Pocket Pocket Head, Chest, Ankle Body	6 17 12 18	no
[30]	2018	Att-DeepConvLSTM	OPPORTUNITY PAMAP2 Skoda	A, G A, G, M A	Body Head, Chest, Ankle Arm	17 12 10	no
Our approach	-	CNN-ResBiGRU-CBAM	WISDM-HARB UTwente	A, G A, G	Wrist Wrist	18 13	yes

3. Methodology

This section presents a sensor-driven framework designed for HAR to achieve recognition accuracy for both symmetric and asymmetric activities. The HAR framework proposed here employs deep learning algorithms to analyze user activities captured by sensors embedded in a wrist-worn wearable device, utilizing signal data collected from these sensors.

To address the challenge of identifying symmetric and asymmetric human activities, we embarked on a study focusing on activity taxonomies [17,18]. We systematically categorized human activities into two primary classes: symmetry and asymmetry. This study aimed to establish precise definitions for symmetric and asymmetric human activities by employing the taxonomy approach.

- Symmetric activities entail the coordinated use of both sides of the body in a mirrored fashion, as depicted in Figure 1a. Common symmetric activities recognized by sensor-based HAR systems include walking, running, climbing stairs, biking, and similar motions.
- In contrast, asymmetric actions involve the body's use in a manner that lacks symmetry or balance, as illustrated in Figure 1b. Rather than exhibiting symmetrical movements across opposing limbs or body parts, these activities feature unpredictable and irregular motions that differ between the sides of the body. Such actions demonstrate unilateral variability, introducing complexity to their analysis. Examples encompass activities like typing, drinking, writing, eating, and other unstructured movements executed with a single hand or on one side of the body, commonly observed daily.

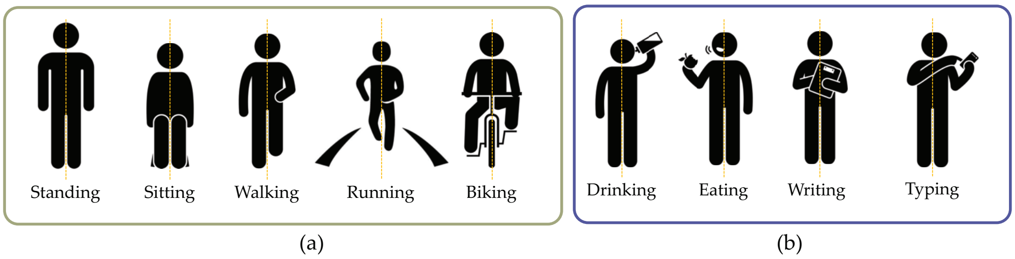


Figure 1. Some samples of different symmetric and asymmetric human activities with respect to the axis of symmetry shown in yellow line: (**a**) symmetric activities; (**b**) asymmetric activities.

3.1. Overview of the Sensor-Based HAR Framework

This section offers a concise overview of the entire setup of the proposed sensor-based HAR framework. The process begins with data acquisition, involving gathering sensor data from wrist-worn devices. Subsequently, the data undergo pre-processing, including noise reduction, handling missing data, and normalization. Data segmentation is then conducted to convert multi-dimensional sensor data into suitable sample data for model training. This involves defining temporal windows, determining their intersections, and assigning classes to segments. The sample data are divided into training and testing sets using 5-fold cross-validation during the data production phase. The subsequent phase entails training various deep learning models, including CNN, LSTM, bidirectional LSTM (BiLSTM), gated recurrent unit (GRU), and bidirectional GRU (BiGRU).

Additionally, we introduce a hybrid residual deep learning model termed CNN-ResBiGRU-CBAM. Performance evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess these models. Figure 2 illustrates the sequential steps of the proposed sensor-based HAR framework.

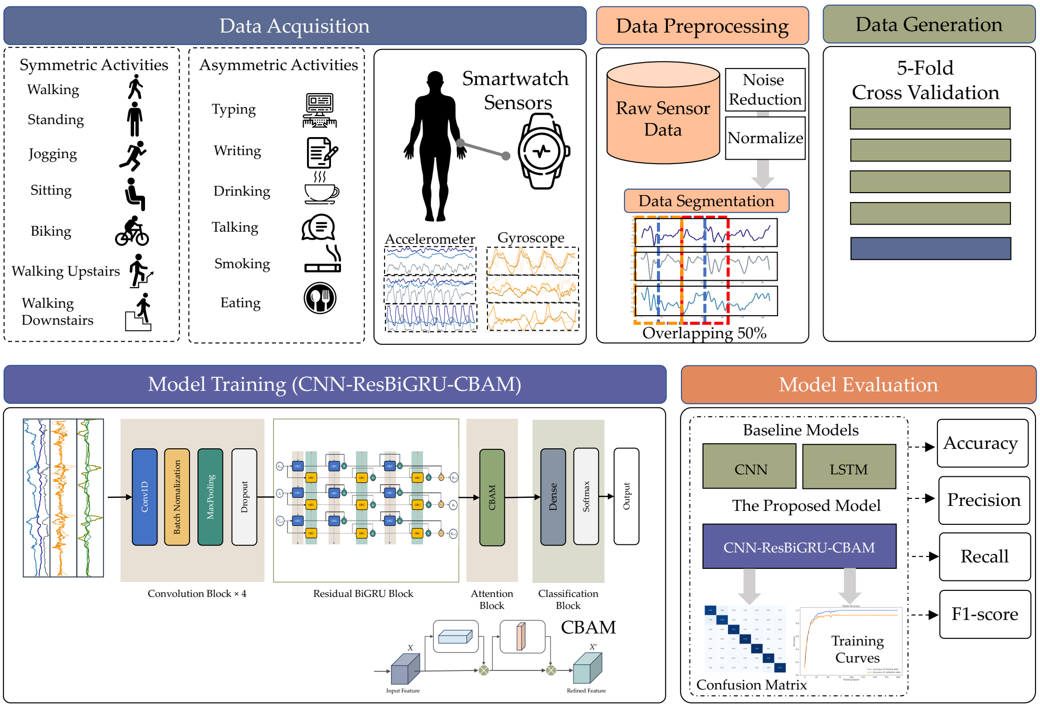


Figure 2. The proposed framework of sensor-based HAR for symmetric and asymmetric human activity recognition.

3.2. Data Acquisition

This study utilizes two publicly accessible benchmark datasets for human activity recognition, encompassing a broad spectrum of symmetric and asymmetric movements recorded through wearable inertial sensors. The first dataset employed is the WISDM Smartphone and Smartwatch Activity and Biometrics dataset, comprising time series data from wearable devices such as triaxial accelerometers and gyroscopes. It encompasses 18 everyday activities, including walking, running, climbing stairs, and sedentary postures. The second dataset utilized is the UTwente dataset, which features data gathered from smartphone and wrist-worn platforms, capturing multimodal sensor data. This dataset encompasses 13 physical activities, ranging from essential to intricate movements, involving sequences of asymmetrical arm movements and whole-body motions. By leveraging these two diverse datasets, which exhibit symmetrical and asymmetrical characteristics, our study facilitates a robust evaluation of our proposed approach in discerning various real-world motion patterns captured by mobile sensors.

3.2.1. WISDM-HARB Dataset

For this study, the widely adopted WISDM-HARB dataset, introduced by Weiss et al. [32], was selected due to its significance and widespread use in identifying human activities through wearable sensors. Given its extensive usage in numerous research endeavors, this dataset is an ideal benchmark for comparing methodologies and evaluating accuracy rates across various implementations. During data collection, the creators employed rigorous supervision and measures to ensure the acquisition of high-quality and consistent movement data. Consequently, the WISDM-HARB dataset is well suited for comprehensive analyses across various activities, particularly excelling in detecting symmetric actions.

The researchers obtained the necessary licenses and approvals from relevant authorities, as the study involved human subjects. Notably, the dataset was publicly available and accessible to all researchers, fostering transparency.

The dataset encompasses a wide array of activities performed by 51 individuals, featuring 18 unique activities, each lasting approximately 3 min. A surveillance setup comprising a wristwatch and a smartphone was employed to monitor the subjects' movements. The wristwatch, worn on the dominant hand, is discreet and compact. Meanwhile, the smartphone is placed in the pocket, mimicking its typical daily carriage. Both devices utilize a custom-built application specifically designed for data collection purposes. Equipped with a total of four sensors, with two sensors allocated to each device, both the smartwatch and smartphone capture accelerometer and gyroscope data. The sampling frequency for each sensor is set at 20 Hz, corresponding to a data sampling every 50 milliseconds. Although a specified data polling rate has been established, it is essential to note that the actual polling rate of sensor data may experience delays if the CPU is occupied.

Symmetric activities are defined by movements that exhibit balance and uniformity on both body sides. The dataset includes numerous symmetric activities, such as walking, running, climbing stairs, sitting, standing, and clapping, as depicted in Table 2. Walking, running, and climbing stairs entail the legs executing repetitive and coordinated motions. Clapping involves repeating rhythmic patterns evenly distributed across both sides of the body's trunk. Sitting and standing represent static, upright postures characterized by symmetrical weight distribution on both body sides. These repetitive, symmetrical movements often demonstrate consistent patterns over time, facilitating activity recognition efforts.

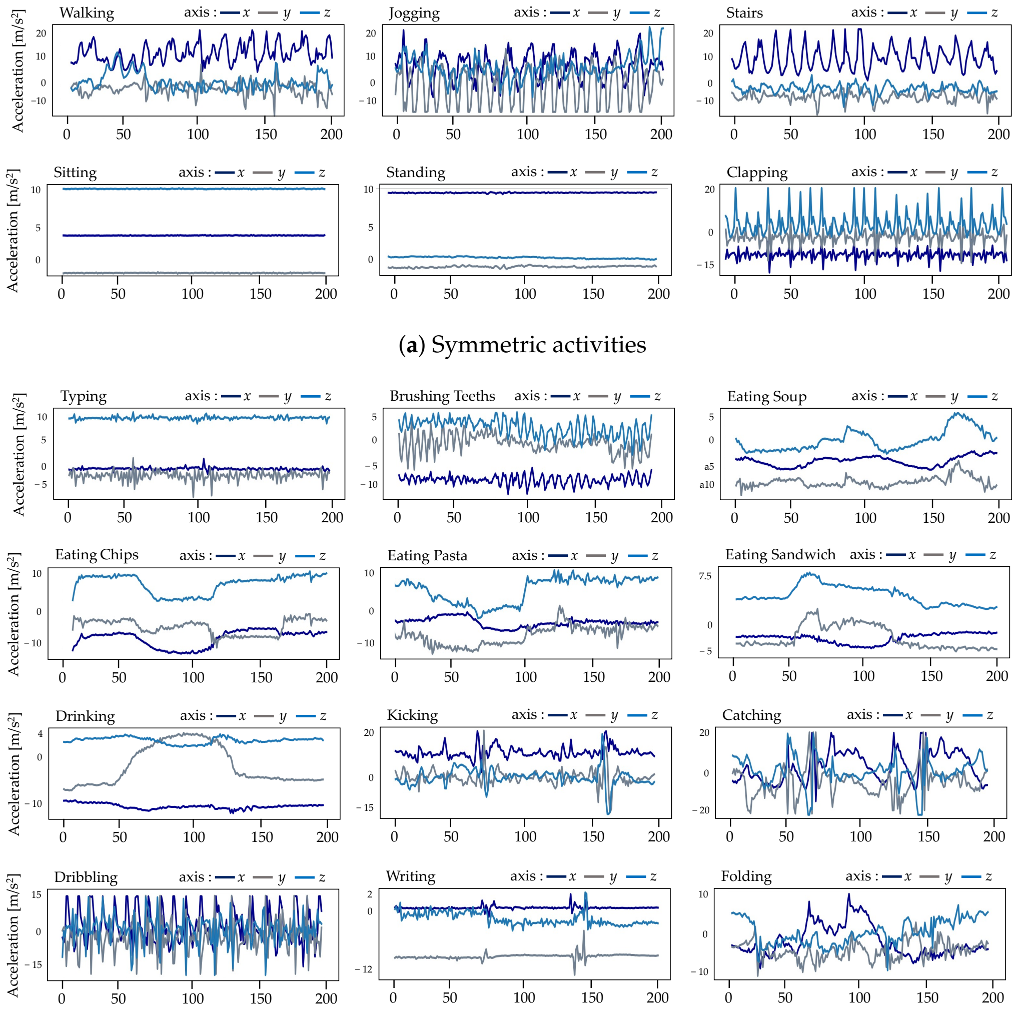
Table 2. Activity list of the WISDM-HARB dataset.

Type	Activity	Description
Symmetic	Walking	Engaging in the activity of moving on foot outside.
	Jogging	Engaging in the activity of running at a steady and moderate pace outside.
	Stairs	Repeatedly ascending and descending many flights of stairs.
	Sitting	Being in a sitting position.
	Standing	Being in an upright position on one's feet.
	Clapping	Striking one's hands together to produce a sound, using both hands.
Asymmetric	Typing	Performing keyboard input tasks while seated.
	Brushing Teeth	Engaging in oral hygiene by brushing teeth.
	Eating Soup	Consuming soup from a bowl.
	Eating Chips	Ingesting snack chips.
	Eating Pasta	Partaking in pasta consumption.
	Eating Sandwich	Consuming a sandwich meal.
	Drinking	Taking liquid refreshment from a cup.
	Kicking	Striking a soccer ball with the foot.
	Catching a ball	Intercepting a thrown object, such as a tennis ball.
	Dribbling	Manipulating a basketball with repeated bounces.
	Writing	Producing written content while seated.
	Folding	Organizing clothing items by creasing and arranging them.

Conversely, asymmetric activities are distinguished by complex and irregular movements that lack distinct symmetrical patterns. Table 2 includes a range of asymmetrical tasks, such as typing, brushing teeth, dribbling a basketball, eating various food items, drinking, and folding clothing. In these activities, movements tend to be less organized spatially, with more significant variability in timing and a lack of consistent rhythm. They typically involve unilateral, asymmetrical movements of specific limbs, often the arms or hands. Differentiating between these disparate and uneven patterns, from both each other and balanced actions, presents significant challenges for accurate identification.

The accelerometer readings capture the forces of acceleration acting on the three sensor axes, including the gravitational force. Upon examining the data samples of symmetric activities (Figure 3a), we observe regular and repetitive patterns that persist over time. These patterns manifest as consistent periodic variations during walking, notable peaks aligning with foot strikes in running, and rhythmic patterns indicative of stair climbing. These patterns correspond to the structured and predictable movements of solid objects. Meanwhile, the gyroscope rotations, illustrated in Figure 4a, display consistent waveforms suggesting dynamic rotational symmetry around the vertical axis during gait activities.

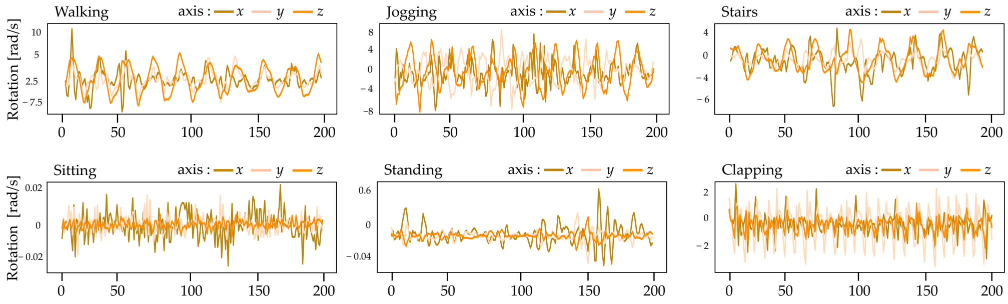
Conversely, the samples of asymmetric activities (Figures 3b and 4b) exhibit a notably higher degree of visual diversity, lacking distinct patterns of repetition. This encompasses sporadic and asymmetrical hand movements during eating, gesturing while brushing, and irregular rotations while handling objects, such as during writing. The fluctuations in accelerations and orientation alterations observed in these intricate activities stem from their spatial and temporal asymmetry.



(**a**) Symmetric activities

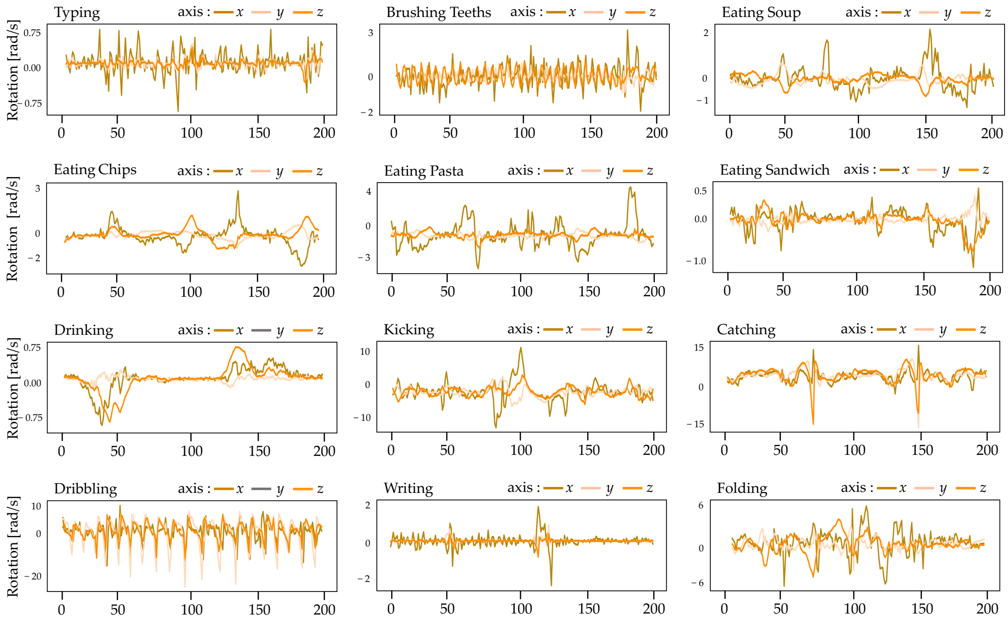
(**b**) Asymmetric activities

Figure 3. Some samples of accelerometer data from the WISDM-HARB dataset.



(**a**) Symmetric activities

Figure 4. *Cont.*



(**b**) Asymmetric activities

Figure 4. Some samples of gyroscope data from the WISDM-HARB dataset.

3.2.2. UTwente Dataset

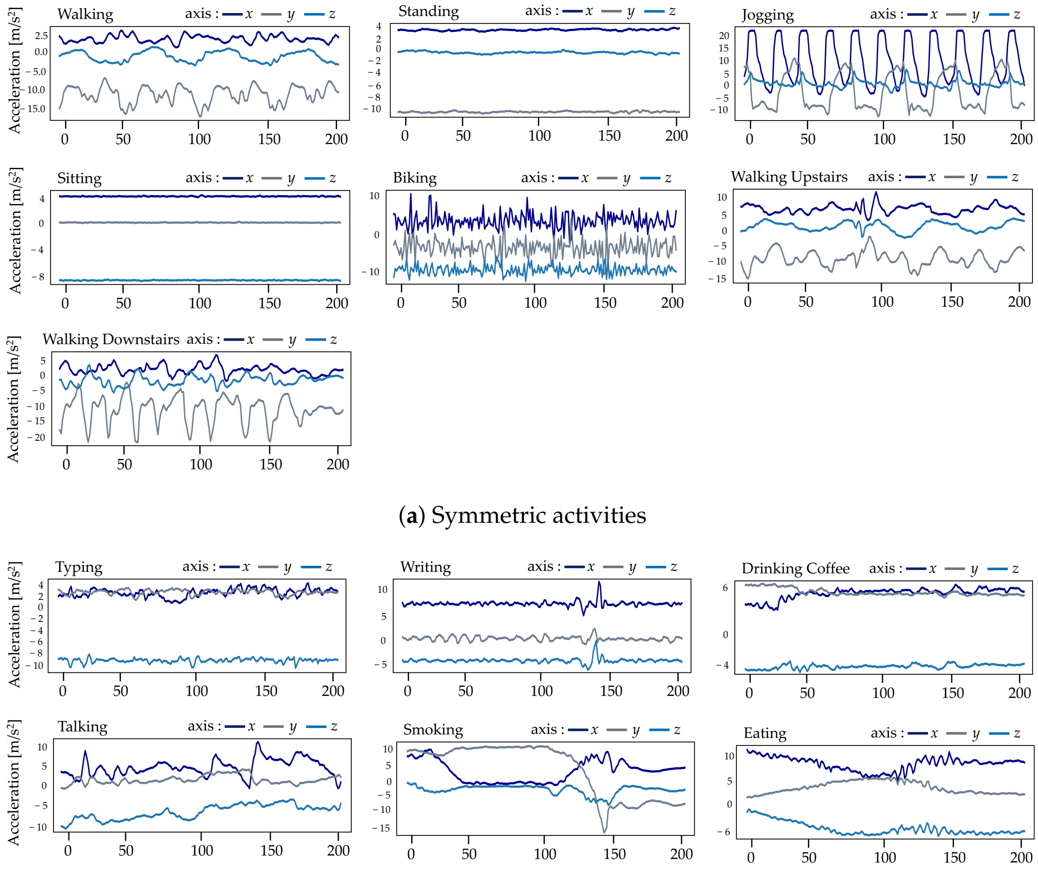
This study analyzes a specific dataset known as the complex human activities using smartphones and smartwatch sensors, commonly called the UTwente dataset [13]. It is a publicly accessible benchmark dataset comprising data from a wrist-worn which, and it was was publicly released by Twente University's pervasive system research group in late 2016. The dataset encompasses data collected from 10 healthy individuals, covering 13 human activities detailed in Table 3. Each participant was instructed to wear two Samsung Galaxy S2 mobile phones, one in their right jeans pocket and the other on their right wrist, mimicking a wristwatch's functionality. Participants engaged in seven fundamental daily tasks to gather sensor-based activity data for three minutes. Additionally, seven participants were tasked with more demanding activities, such as eating, typing, writing, drinking coffee, and conversing, for 5–6 min. Six out of the ten subjects were smokers and were instructed to smoke a single cigarette. To ensure a balanced class distribution, the authors utilized 30 min of data from every participant for each activity. Data were collected from an accelerometer, a linear acceleration sensor, a gyroscope, and a magnetometer at a frequency of 50 Hz.

Table 3. Activity list of the UTwente dataset.

Type	Activity	Description
Symmetic	Walking Jogging Standing Sitting Biking Walking Upstairs Walking Downstairs	Walking at a normal pace on a flat surface indoors Jogging at a moderate pace on a flat surface indoors Standing still in an upright position Sitting in a chair with minimal movement Riding a bicycle outdoors on a flat surface Climbing multiple flights of stairs in an upward direction Descending multiple flights of stairs in a downward direction
Asymmetric	Typing Writing Drinking Coffee Talking Smoking Eating	Typing on a computer keyboard while seated in a chair Handwriting with a pen on paper while seated in a chair Consuming a beverage from a cup while seated Engaging in a conversation while standing still Smoking a cigarette while standing still Consuming a cup of soup using a spoon while seated

The UTwente dataset consists of a range of typical daily activities that exhibit symmetrical and asymmetrical motion characteristics. The symmetric activities involve consistent and repetitive movement patterns, such as walking, jogging, cycling, and climbing stairs. Stationary postures like standing and sitting are also considered symmetric. In contrast, asymmetric activities involve more complex and irregular movements, primarily using the hands and arms. These include typing on a keyboard, handwriting, drinking, smoking, engaging in conversation, and eating with utensils. The dataset comprehensively represents various human activities in a realistic setting, captured using inertial sensors on smartphones and smartwatches.

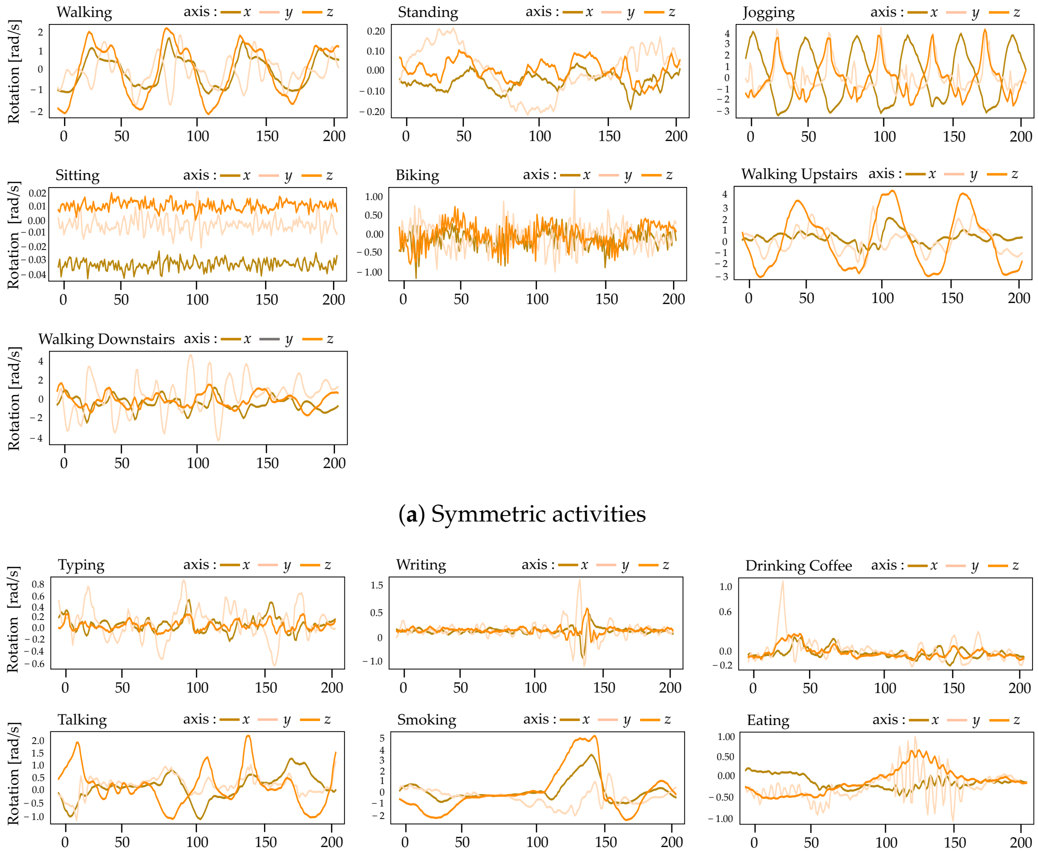
The accelerometer data exhibit recognizable repetitive patterns during symmetric activities, such as the consistent cyclic fluctuations observed during walking and jogging (as illustrated in Figure 5a). In contrast, asymmetric activities like smoking, talking, and eating produce irregular and fluctuating signals (as depicted in Figure 5b). Similarly, the gyroscope's rotations exhibit structured patterns over time during activities like cycling with regular leg pedaling (as seen in Figure 6a), in contrast to the unpredictable orientations observed during asymmetrical actions like typing, writing, or drinking (as shown in Figure 6b). Both sensor feeds present distinct patterns over time, providing clear evidence of the symmetry and asymmetry inherent in the observed complex human behaviors.



(**a**) Symmetric activities

(**b**) Asymmetric activities

Figure 5. Some samples of accelerometer data from the UTwente dataset.



(**a**) Symmetric activities

(**b**) Asymmetric activities

Figure 6. Some samples of gyroscope data from the UTwente dataset.

3.3. Data Pre-Processing

Pre-processing is essential to read the raw inertial sensor data for training algorithms that recognize activities. The publicly available datasets comprise multi-dimensional time series data gathered from wrist-worn devices worn by participants, encompassing the 3-axis accelerometer and gyroscope recordings. This study implemented a systematic data preparation pipeline before constructing deep neural networks. The pipeline involved denoising, normalizing, and segmenting the data.

3.3.1. Data Denoising

When handling time series data, it is typically essential to employ noise-reduction methods. Sensor measurements are often subject to uncertainties, which can manifest as noise in the signal. Complex tasks involving successive movements contribute to noise accumulation in data recordings from inertial sensors. In this study, noise reduction was achieved by applying a median filter and a third-order low-pass Butterworth filter, with a cutoff frequency set at 20 Hz. This frequency was sufficient to capture bodily motions, as the energy content below 15 Hz accounted for 99% of the signal [33].

3.3.2. Data Normalization

Equation (1) demonstrates the scaling of the initial sensor data to a standardized interval from 0 to 1. This method simplifies the model's learning task by ensuring all data points fall within a limited range. Consequently, gradient descents can achieve quicker convergence rates.

$$x_i^{norm} = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}, i = 1, 2, 3, ... \tag{1}$$

where x_i^{norm} represents the normalized data, n represents the number of channels, and x_i^{max} and x_i^{min} are the maximum and minimum values of the i-th channel, respectively.

3.3.3. Data Segmentation

The wearable devices accumulate a substantial volume of sensor data. Inputting the complete dataset into the HAR model simultaneously is impractical. To address this issue, we utilize the sliding window technique. This method segments data and finds extensive usage in HAR systems [34]. It caters to symmetric activities like running, walking, and standing, as well as asymmetric activities such as drinking, eating, and writing [35].

The sliding window method entails splitting the unprocessed sensor readings into segments of predetermined length, with a specified amount of overlap between adjoining segments. This approach augments the number of samples for training purposes and effectively captures transitions between different activities. For this study, we divided the sensor data into 10-s segments, with a 50% overlap between consecutive segments.

Figure 7 depicts segmenting data using the sliding window method. The sensor readings are split into segments with a fixed 10-s width. Each segment overlaps with the preceding one by 50%. This overlapping allows for a more seamless transition between consecutive segments. It also aids in effectively capturing the interdependencies that exist across sequential data points over time.

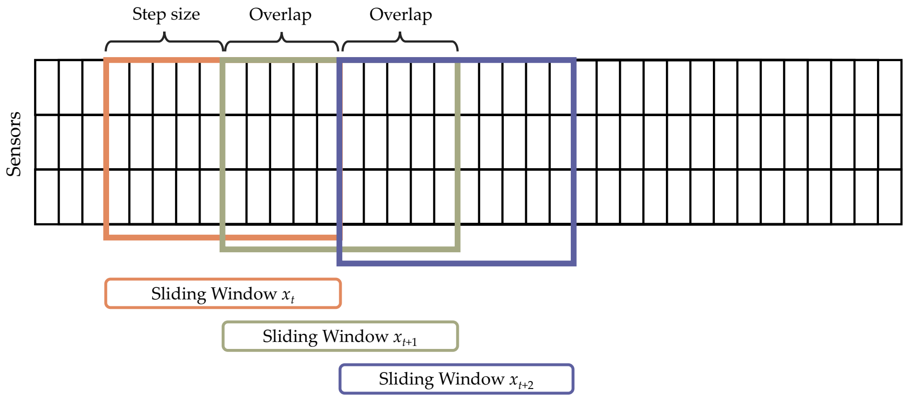


Figure 7. A visualization of the sliding window segmentation with a fixed-width window size of 10 s with the overlapping of 50%.

Utilizing the sliding window technique, we convert the uninterrupted sensor readings into a sequence of segments with predetermined lengths. The HAR model can efficiently handle these segmented inputs. This data segmentation approach allows the model to discern the unique patterns and characteristics of various human activities, enhancing the accuracy of activity identification.

3.4. The Proposed CNN-ResBiGRU-CBAM Model

The suggested model comprises three main elements: convolution, residual BiGRU, and CBAM blocks. It represents an end-to-end deep learning framework, functioning as a hybrid architecture. The overall layout of the proposed model is depicted in Figure 8.

The first module, the convolution block (ConvB), is responsible for extracting spatial features from the pre-processed information. Modifying the convolution kernel's stride may significantly decrease the temporal series' length, hence improving the recognition speed. After completing this stage, the BiGRU network extracts temporal characteristics from the data that the convolution block has processed. This component improves the model's ability to capture long-term dependencies in time series data using the characteristics of a BiGRU. This integration improves the model's understanding of complex temporal patterns and strengthens its recognition accuracy. To further enhance the final recognition qualities, we used an attention mechanism called a convolutional block attention module (CBAM). This technique calculates weights for the feature maps generated by the BiGRU network, allowing the model to concentrate on the most exciting parts of the input data. CBAM increases the model's discriminatory capacity and boosts activity recognition accuracy by emphasizing the most significant characteristics. Ultimately, the behavior information is classified by the wholly linked layer and SoftMax function. The result of this categorization procedure serves as the recognition output, predicting the particular action in progress. The following sections will comprehensively explain each component, clearly outlining their functions and contributions within our suggested model.

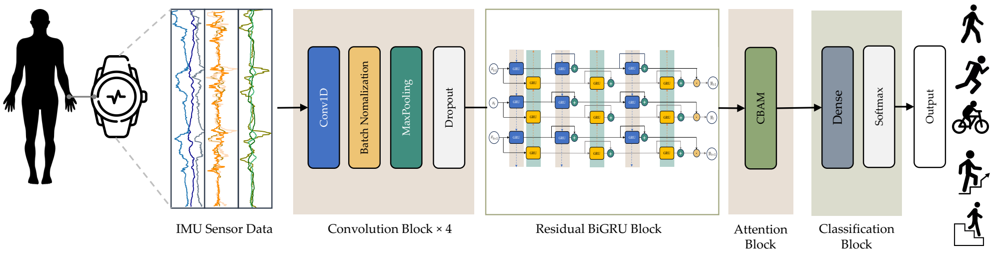


Figure 8. Detailed and architecture of the proposed CNN-ResBiGRU-CBAM model.

3.4.1. Convolution Block

When a CNN is deployed, it usually relies on a predetermined collection of components. CNNs are often used in the context of supervised learning. Usually, these neural networks establish connections between each neuron and every other neuron in the subsequent network layers. The activation function of a neural network transforms the input value of neurons into the corresponding output value. Two critical factors influence the efficiency of the activation function. These factors include the scarcity of data and the ability of the neural network's lower layers to withstand the reduced flow of gradients. CNNs often use pooling to reduce the dimensionality of data. The maximum and average pooling processes are often used and referred to as max-pooling and average pooling, respectively.

The work uses ConvB to extract basic features from unprocessed sensor data. Figure 8 illustrates that ConvB consists of four layers: 1D-convolutional (Conv1D) and batch normalization (BN), the max-pooling layer (MP), and the dropout layer. The Conv1D utilizes several trainable convolutional kernels to capture different attributes, with each kernel generating a unique feature map. The decision was made to use the BN layer to enhance stability and speed up the training process.

3.4.2. Residual BiGRU Block

Human actions are time-based; therefore, it depends on whether the convolution block for extracting spatial features is inadequate for recognizing actions. It is necessary to consider the chronological order of the whole event. RNNs have advantageous skills for handling time series data. Nevertheless, RNN models can encounter gradient vanishing and data loss when the time series expands.

Hochreiter et al. [36] introduced an LSTM network. LSTM, as opposed to basic RNNs, is a recurrent neural network that uses gates to successfully store and remember information over extended periods. Furthermore, it surpasses ordinary RNNs in effectively managing more extended time series. However, behavioral data are impacted not only by prior times but also by following events.

While LSTM has successfully addressed the vanishing gradient in RNNs, its memory cells contribute to higher memory use. In 2014, Cho et al. [37] proposed the GRU network, a new model based on RNNs. The GRU is a simplified variant of the LSTM that lacks a distinct memory cell in its architecture [38]. Within a GRU network, updates and reset gates govern the extent to which each hidden state is modified. It discerns which information should be transmitted to the subsequent stage and which should not. A BiLSTM is a neural network architecture incorporating forward and backward information using two GRU networks. BiGRU improves time series feature extraction by capturing bidirectional relationships, in contrast to the GRU network. Hence, using a BiGRU network to extract time series characteristics from behavioral data is a suitable methodology.

The BiLSTM network excels in capturing time series features but lacks spatial information gathering. Increasing the number of stacked layers can lead to gradient vanishing during training. To tackle this problem, the Microsoft Research team introduced ResNet [39] in 2015, achieving a depth of 152 layers and winning the ILSVRC competition. The formulation for each residual block of the ResNet architecture is provided as follows:

$$x^{i+1} = x^i + f(x^i, W_i) \tag{2}$$

The residual blocks are partitioned into two components: x^i, which represents a direct mapping, $F(x^i, W_i)$, representing the residual portion.

$$\hat{x}^i = \frac{x^i - E(x^i)}{\sqrt{var(x^i)}} \tag{3}$$

where x^i denotes the input vector in the i-th dimension and $\hat{x}^i$ represents the output subsequent to layer normalization.

This research introduces a novel combination of residual structure and layer normalization in a BiGRU network, referred to as ResBiGRU. The diagram illustrating this combination can be seen in Figure 9. The recursive feature information y could be defined as

$$x_t^{f(i+1)} = LN(x_t^{f(i)} + G(x_t^{f(i)}, W_i)) \tag{4}$$

$$x_t^{b(i+1)} = LN(x_t^{b(i)} + G(x_t^{b(i)}, W_i)) \tag{5}$$

$$y_t = concat(x_t^f, x_t^b) \tag{6}$$

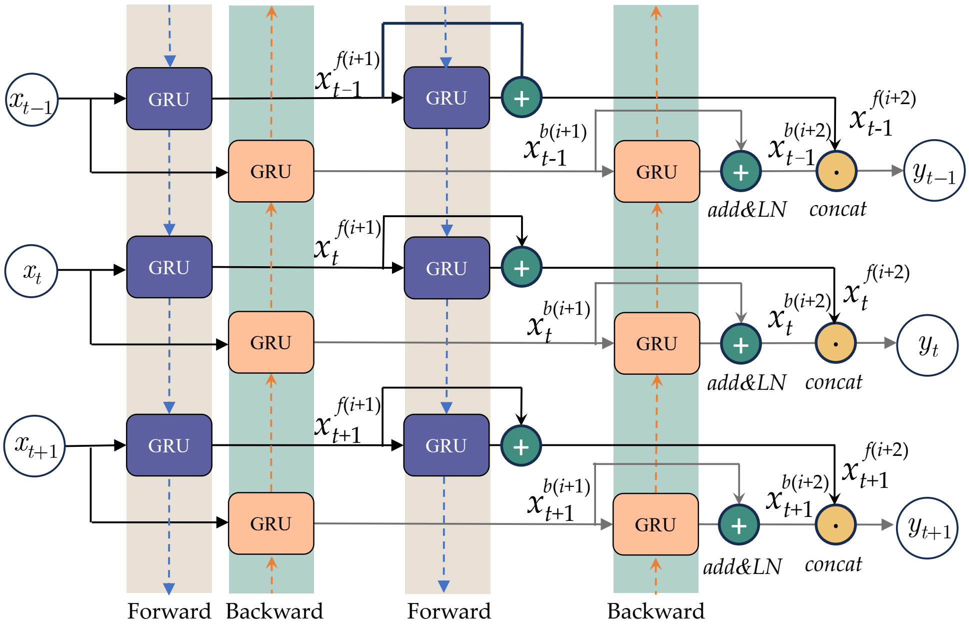


Figure 9. Structure of the ResBiGRU.

Layer normalization (LN) denotes the normalization of layers. At the same time, the processing of input states in the GRU network are represented by G. The t-th moment in the time series is denoted by the subscript t in $x_t^{f(i+1)}$. The forward state is indicated by the superscript f, the reverse state by b, and the number of stacked layers by $(i+1)$. The encoded information y_t at time t is generated by amalgamating the forward and backward states.

3.4.3. CBAM Block

To augment feature representation and elevate HAR performance, the proposed deep residual network integrates the convolutional block attention module (CBAM) [40]. CBAM adaptively refines the feature maps by sequentially applying attention mechanisms focused on channels and spatial attributes. The architectural structure of CBAM is depicted in Figure 10.

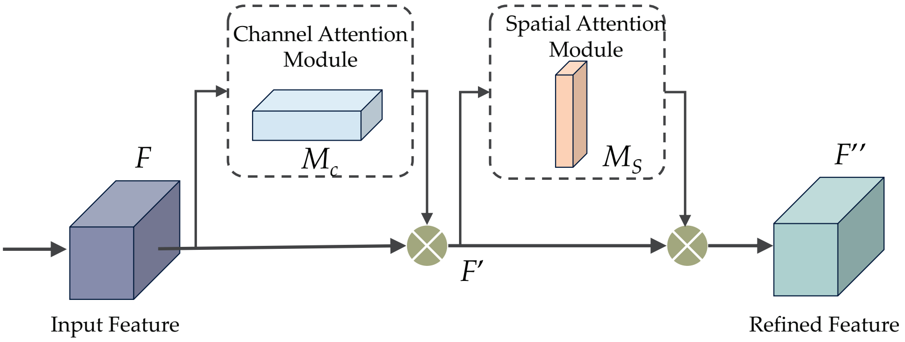


Figure 10. The convolutional block attention module (CBAM).

The channel attention module leverages the interdependencies among different feature channels, while the spatial attention module concentrates on the interrelationships across spatial locations. By applying these attention mechanisms, CBAM enables the network to focus on more informative features and suppress less relevant ones.

Given an input feature map, the channel attention module first generates a channel attention map by employing maximum and average pooling operations along the spatial dimensions. These pooled features are processed through a shared multi-layer perceptron (MLP) to produce the channel attention weights. On the other hand, the spatial attention module generates a spatial attention map by applying maximum and average pooling operations along the channel dimension, followed by a convolutional layer to produce the spatial attention weights.

The refined feature map is obtained by linearly multiplying the input feature map with the channel and spatial attention weights. This allows the network to adjust the importance of different channels and spatial locations adaptively [41], thereby enhancing the discriminative power of the learned features for HAR tasks.

3.5. CNN-ResBiGRU-CBAM Model Hyperparameters

Table 4 briefly summarizes the parameters employed in the research concerning the CNN-ResBiGRU-CBAM framework. This model comprises four key phases: the convolutional block, the residual BiGRU block, the CBAM block, and the classification block.

The convolutional segment undergoes four iterations and comprises Conv1D layers utilizing a kernel size of 3, a stride of 1, and 256 filters. ReLU activation, batch normalization, max pooling, and dropout procedures follow these layers. These layers extract and analyze specific patterns and traits within the input data. Next, the residual BiGRU portion consists of two BiGRU layers, one with 128 neurons and the other with 64 neurons. These layers capture temporal dependencies and contextual information over extended periods. The CBAM segment enhances feature representations by selectively focusing on crucial channels and spatial locations. Lastly, the classification component includes a densely connected layer with a neuron count matching the number of activity classes. The application of the SoftMax activation function follows this. SoftMax converts the dense layer's output values into a probability distribution across the activity classes. Formally, it can be defined as follows:

$$\text{SoftMax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \tag{7}$$

Assume x_i represents the input to the i-th neuron within the dense layer, and let K indicate the overall count of activity categories. SoftMax ensures that the resulting probabilities sum up to 1, rendering it suitable for tasks entailing the classification of numerous categories [42].

During training, the model employs the cross-entropy loss function to assess the variance between the predicted probability and the actual activity labels. This loss function, cross-entropy, is a mathematical expression utilized for gauging the distinction between two probability distributions, which is defined as

$$\text{Loss} = -\sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij} \log(\hat{y}_{ij}) \tag{8}$$

Consider N as the count of samples, K as the count of activity categories, y_{ij} as the genuine label (either 0 or 1) for the i-th sample regarding the j-th category, and $\hat{y}_{ij}$ as the estimated probability of the i-th sample for the j-th category. The cross-entropy loss function is frequently utilized in deep learning for classification tasks and has demonstrated favorable performance and convergence characteristics [43].

To enhance the model, the Adam optimizer is employed, which adjusts the learning rate for each parameter based on its previous gradients [44]. Training is executed using a batch size of 128 and over 200 epochs to empower the model to grasp robust feature representations and achieve high generalization capability.

Table 4. The summary of hyperparameter s of the CNN-ResBiGRU-CBAM used in this work.

Stage	Hyperparameters		Values
Architecture	Convolutional Block × 4		
	1D Convolution	Kernel Size	3
		Stride	1
		Filters	256
		Activation	ReLU
	Batch Normalization		-
	Max Pooling		2
	Dropout		0.25
	Residual BiGRU Block		
	ResBiGRU_1	Neural	128
	ResBiGRU_2	Neural	64
	CBAM Block		
	CBAM Layer		-
	Classification Block		
	Dense		Number of activity classes
	Activation		SoftMax
Training	Loss Function		Cross-entropy
	Optimizer		Adam
	Batch Size		128
	Number of Epochs		200

3.6. Cross-Validation

In order to evaluate the effectiveness of the CNN-ResBiGRU-CBAM model, we used the k-fold cross-validation (k-CV) technique [45]. This method entails dividing the dataset into k subsets that are nearly equal in size, different from each other, and non-overlapping. After the data are divided into subsets, one is chosen as the validation set, while the other $k-1$ subsets are used to train the model. The total performance is calculated using the average performance parameters, such as accuracy, precision, recall, and F1-score, over all k folds [46].

It is essential to highlight that the k-CV approach can require considerable computational resources, particularly with large datasets or when using high k values. Utilizing the k-CV technique aims to guarantee a just and impartial assessment of the model [47]. In our investigation, we opted for a 5-fold cross-validation ($k = 5$) to find a compromise between computational efficiency and accurate performance estimation.

4. Experiments and Results

In this section, we offer a thorough evaluation to determine the efficiency of the CNN-ResBiGRU-CBAM method. We demonstrated the effectiveness of this approach on two standard HAR datasets (WISDM-HARB and UTwente), comparing it against various baseline deep learning architectures like CNN, LSTM, BiLSTM, GRU, and BiGRU. To gauge the performance of the DL models in SAR applications, we relied on accuracy and the F1-score, which are widely recognized metrics.

4.1. Experimental Settings

This research leveraged Google Colab Pro+ in conjunction with a Tesla V100-SXM2-16GB, Hewlett Packard Enterprise, Los Angeles, CA, USA, graphics processing unit to accelerate the training of deep learning models. The CNN-ResBiGRU-CBAM and other foundational deep learning architectures were implemented using a Python 3.6.9 framework that employs TensorFlow 2.2.0 and CUDA 10.2 backends. The investigation focused on utilizing several Python libraries, including

- Numpy and Pandas manage data during data retrieval, processing, and sensor data analysis.
- Matplotlib and Seaborn are used to craft visualizations and present the results of data analysis and model evaluation.
- Scikit-learn, or Sklearn, is used to gather and generate data for research endeavors.
- TensorFlow is used to construct and train deep learning models.

Multiple tests were performed on the WISDM-HARB and UTwente datasets to assess the most effective method. The trials used a five-fold cross-validation process.

4.2. Experimental Results

In this section, we conduct thorough comparative evaluations of two widely recognized benchmark datasets, namely WISDM-HARB and UTwente. We aim to assess the effectiveness of our proposed deep CNN design, CNN-ResBiGRU-CBAM, for sensor-based HAR. Through extensive empirical investigations, we demonstrate the learning ability to adeptly discern unique spatial and temporal patterns from data captured by inertial sensors. These capabilities enable the accurate classification of both typical symmetric motions and complex asymmetric activities.

4.2.1. Experimental Results from the WISDM-HARB Dataset

We outline the evaluation experiments and the activity recognition results achieved using the WISDM-HARB benchmark dataset. Our objective is to validate the effectiveness of the CNN-ResBiGRU-CBAM model architecture proposed in our study. Using this widely utilized public dataset, we assess the performance of various foundational deep learning models, such as CNN, LSTM, BiLSTM, GRU, and BiGRU.

Based on the data presented in Table 5, which illustrates the accuracy of deep learning models in recognizing patterns in accelerometer data from the WISDM-HARB dataset, several key observations can be made. The CNN model achieves a modest accuracy of 69.10% for identifying human activities, indicating that relying solely on convolutional layers for spatial feature extraction may not adequately capture the temporal dynamics of human movements. In contrast, recurrent models such as LSTM, BiLSTM, GRU, and BiGRU exhibit notable improvements in accuracy, ranging from 81% to 85%, owing to their ability to model temporal sequences effectively. This underscores the importance of integrating temporal modeling alongside spatial feature learning. Specifically, BiLSTM and BiGRU models, which analyze sequences bidirectionally, outperform their unidirectional counterparts (LSTM and GRU), highlighting the benefits of incorporating past and future contexts for activity classification. Our CNN-ResBiGRU-CBAM model achieves an impressive accuracy of 86.77%, surpassing the performance of all other deep learning models used for comparison. This underscores the effectiveness of our architectural enhancements.

Table 5. Recognition performance of deep learning models using accelerometer data of the WISDM-HARB dataset.

Model	Recognition Performance			
	Accuracy	Precision	Recall	F1-Score
CNN	69.10%	68.77%	69.11%	68.62%
LSTM	81.08%	81.13%	81.07%	80.94%
BiLSTM	84.14%	84.17%	84.13%	84.05%
GRU	81.39%	81.41%	81.38%	81.25%
BiGRU	85.39%	85.43%	85.39%	85.36%
CNN-ResBiGRU-CBAM	86.77%	86.90%	86.77%	86.69%

The recognition performance of gyroscope data from the WISDM-HARB dataset is depicted in Table 6. Similar patterns in performance are observed compared to the accelerometer data. Once again, the CNN model demonstrates the lowest accuracy rate of 59.36%, underscoring the limitations of relying solely on spatial feature learning for activity recognition. In contrast, the LSTM, BiLSTM, GRU, and BiGRU recurrent networks achieve higher accuracy rates ranging from 71% to 73%, highlighting their effectiveness in capturing temporal sequences. The BiLSTM and BiGRU models outperform the LSTM and GRU models, indicating the benefits of incorporating bidirectional context modeling for gyroscope-based activity classification. Our proposed CNN-ResBiGRU-CBAM architecture attains an accuracy of 75.13%, further validating the advantages of integrating CNN, RNN, and attention mechanisms. Since the gyroscope measures orientation and rotational movements, accurately modeling temporal changes is crucial for identifying gesture-based movements.

Table 6. Recognition performance of deep learning models using the gyroscope data of the WISDM-HARB dataset.

Model	Recognition Performance			
	Accuracy	Precision	Recall	F1-Score
CNN	59.36%	59.08%	59.31%	58.90%
LSTM	73.34%	73.17%	73.33%	73.81%
BiLSTM	73.89%	73.66%	73.56%	73.84%
GRU	71.21%	72.76%	71.20%	71.50%
BiGRU	72.34%	72.26%	72.31%	72.19%
CNN-ResBiGRU-CBAM	75.13%	75.28%	75.12%	73.76%

Upon comparing the performance of deep learning models on both accelerometer (Table 5) and gyroscope (Table 6) data from the WISDM-HARB dataset, certain conclusions emerge. The accuracy of all models experiences a notable enhancement when utilizing combined accelerometer and gyroscope data, as evidenced in Table 7, compared to using either sensor in isolation. This investigation corroborates the notion that amalgamating diverse motion modalities improves accuracy in identifying activities. For instance, the accuracy of the CNN model increases from 69.10% with individual sensors to 72.27% with multimodal data. Similarly, significant enhancements are observed for LSTM (82.63% compared to 81.08%), BiLSTM (86.00% compared to 84.14%), GRU (84.77% compared to 81.39%), and BiGRU (86.92% compared to 85.39%). When employing sensor fusion, our proposed CNN-ResBiGRU-CBAM architecture achieves an accuracy of 89.01%, surpassing the accuracies achieved using accelerometer and gyroscope signals separately, which were 86.77% and 75.13%, respectively.

Table 7. Recognition performance of deep learning models using both the accelerometer and the gyroscope data of the WISDM-HARB dataset.

Model	**Recognition Performance**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
CNN	72.27%	72.38%	72.24%	71.96%
LSTM	82.63%	82.55%	82.61%	82.49%
BiLSTM	86.00%	86.01%	85.99%	85.93%
GRU	84.77%	84.85%	84.78%	84.70%
BiGRU	86.92%	86.88%	86.92%	86.83%
CNN-ResBiGRU-CBAM	89.01%	89.00%	89.01%	88.94%

4.2.2. Experimental Results from the UTwente Dataset

Detailed comparative evaluations are provided in this section, focusing on the UTwente benchmark dataset to extend the investigation into the performance of our proposed CNN-ResBiGRU-CBAM architecture in recognizing human activities using sensor data.

An analysis of the recognition performance data displayed in Table 8, focusing on accelerometer data from the UTwente dataset, yields significant insights. The CNN model exhibits an accuracy of 85.55%, outperforming its performance on the WISDM-HARB dataset. This indicates a more efficient process of learning spatial features in UTwente activities. While the LSTM and GRU models achieve accuracy rates of 84.70% and 93.55%, respectively, they lag behind the BiLSTM (90.64%) and BiGRU (95.68%) models, highlighting the advantages of bidirectional context modeling. Our proposed CNN-ResBiGRU-CBAM architecture achieves an impressive accuracy of 96.15%, underscoring the effectiveness of our residual connections and attention modules.

Table 8. Recognition performance of deep learning models using the accelerometer data of the UTwente dataset.

Model	**Recognition Performance**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
CNN	85.55%	86.28%	85.53%	85.13%
LSTM	84.70%	84.75%	84.70%	83.90%
BiLSTM	90.64%	91.02%	90.64%	90.51%
GRU	93.55%	93.80%	93.55%	93.51%
BiGRU	95.68%	95.73%	95.68%	95.65%
CNN-ResBiGRU-CBAM	96.15%	96.39%	96.15%	96.14%

Examination of the gyroscope data from the UTwente dataset, as shown in Table 9, reveals the following comparative trends in the recognition performance of deep learning models: The CNN model achieves a decreased accuracy of 72.08%, indicating that spatial features alone provide less insight for the gyroscope compared to the accelerometer. The LSTM and BiLSTM models also demonstrate significantly reduced accuracy, at 38.86% and 64.12%, respectively, indicating challenges in capturing complex rotational motions effectively. Surprisingly, the GRU model (81.53%) outperforms the BiGRU model (75.17%), in contrast to observations on the accelerometer data, suggesting that unidirectional context holds more significance for UTwente gyroscope signals. Our proposed CNN-ResBiGRU-CBAM model achieves an accuracy of 88.93%, notably higher than previous state-of-the-art models, affirming its robustness.

Table 9. Recognition performance of deep learning models using the gyroscope data of the UTwente dataset.

Model	Recognition Performance			
	Accuracy	Precision	Recall	F1-Score
CNN	72.08%	72.44%	72.08%	71.72%
LSTM	38.86%	38.52%	38.86%	35.32%
BiLSTM	64.12%	63.92%	64.13%	62.23%
GRU	81.53%	81.93%	81.53%	81.20%
BiGRU	75.17%	75.36%	75.17%	74.34%
CNN-ResBiGRU-CBAM	88.93%	89.79%	88.93%	88.45%

A thorough examination of multimodal recognition performance utilizing combined accelerometer and gyroscope signals from the UTwente dataset, outlined in Table 10, reveals several noteworthy observations. The integration of data from both the accelerometer and gyroscope significantly boosts the accuracy of all models, confirming the significance of incorporating supplementary information encoding human movements. Specifically, the CNN's accuracy improves from 85.55% and 72.08% when utilizing solely accelerometer and gyroscope data to 93.07% when amalgamating data from both sensors through sensor fusion. Consistent enhancements are observed across LSTM, BiLSTM, GRU, and BiGRU models, with BiLSTM's accuracy rising from 90.64% and 64.12% to 93.29%. This underscores the advantages of fusion in capturing temporal sequences. Our proposed CNN-ResBiGRU-CBAM architecture achieves an accuracy of 96.49% when leveraging multimodal data, surpassing the accuracies achieved from the individual accelerometers and gyroscope signals, which were 96.15% and 88.93%, respectively.

Table 10. Recognition performance of deep learning models using both accelerometer and gyroscope data of UTwente dataset.

Model	Recognition Performance			
	Accuracy	Precision	Recall	F1-Score
CNN	93.07%	93.22%	93.07%	92.95%
LSTM	90.00%	90.53%	90.00%	89.82%
BiLSTM	93.29%	93.42%	93.28%	93.25%
GRU	94.66%	94.93%	94.66%	94.60%
BiGRU	95.72%	95.80%	95.72%	95.71%
CNN-ResBiGRU-CBAM	96.49%	96.62%	96.49%	96.47%

Integrating complementary data on acceleration and rotation enables a more comprehensive examination of HAR. Our approach adeptly manages varied sequences by incorporating unified feature extraction and bidirectional context modeling elements.

4.3. Comparison with State-of-the-Art Models

Furthermore, we carried out a comparative analysis of the CNN-ResBiGRU-CBAM model proposed in this study against other cutting-edge deep learning models known for achieving superior results in HAR. The first contender is the InceptionTime model, as proposed in [48], which integrates a modified Inception architecture with a gated recurrent unit and residual connections to enhance recognition accuracy, particularly on imbalanced HAR datasets. The second competitor is the DeepConvTCN model [49]. This deep learning architecture comprises a deep convolutional neural network (DeepConv) and temporal convolutional networks (TCNs). Within this research, we assessed the performance of our CNN-ResBiGRU-CBAM model by comparing it to these two deep learning models using the WISDM-HARB and UTwente benchmark HAR datasets.

Using the WISDM-HARB dataset, the proposed CNN-ResBiGRU-CBAM model outperformed the DeepConvTCN and InceptionTime models, as shown in Table 11. Symmetric activities achieved the highest F1 scores in four out of six activities, with InceptionTime being better for sitting and standing. The proposed model obtained the top F1-scores for asymmetric activities in six out of nine activities, with DeepConvTCN and InceptionTime each leading in one activity.

Table 11. Comparison of F1-score between the proposed CNN-ResBiGRU-CBAM model and state-of-the-art models using wearable sensor data from the WISDM-HARB dataset.

Type	Activity	F1-Score		
		DeepConvTCN [49]	InceptionTime [48]	The Proposed CNN-ResBiGRU-CBAM
Symmetric	Walking	0.96	0.93	0.97
	Jogging	0.98	0.73	1.00
	Stairs	0.88	0.98	0.94
	Sitting	0.78	0.80	0.80
	Standing	0.82	0.91	0.86
	Clapping	0.98	0.86	0.97
Asymmetric	Typing	0.88	0.80	0.92
	Brushing Teeth	0.98	0.98	0.96
	Eating Soup	0.86	0.87	0.85
	Eating Chips	0.75	0.73	0.70
	Eating Pasta	0.82	0.83	0.84
	Drinking	0.87	0.86	0.82
	Eating Sandwishes	0.62	0.73	0.58
	Kicking	0.90	0.81	0.95
	Catching a ball	0.93	0.90	0.97
	Dribbling	0.94	0.90	0.98
	Writing	0.86	0.72	0.94
	Folding	0.87	0.88	0.97
	Average	0.87	0.85	0.89

Overall, the CNN-ResBiGRU-CBAM model attained the highest average F1-score of 0.89 across all activities, surpassing DeepConvTCN (0.87) and InceptionTime (0.85). This demonstrates its superior performance and robustness in recognizing symmetric and asymmetric activities using the WISDM-HARB wearable sensor dataset.

The proposed CNN-ResBiGRU-CBAM model outperformed the DeepConvTCN and InceptionTime models in activity recognition using the UTwente dataset, as shown in Table 12. It achieved the highest F1 scores for symmetric activities in six out of seven activities, with InceptionTime performing better only for sitting. The proposed model obtained the highest scores for asymmetric activities in five out of six activities, with InceptionTime slightly better for eating.

Overall, the CNN-ResBiGRU-CBAM model attained the highest average F1 score of 0.965 across all activities, surpassing DeepConvTCN (0.918) and InceptionTime (0.922). This demonstrates its superior performance and generalization ability in recognizing both symmetric and asymmetric activities using wearable sensor data from the UTwente dataset.

Table 12. Comparison of F1 score between the proposed CNN-ResBiGRU-CBAM model and state-of-the-art models using wearable sensor data from the UTwente dataset.

Type	Activity	F1-Score		
		DeepConvTCN [49]	InceptionTime [48]	The Proposed CNN-ResBiGRU-CBAM
Symmetric	Walking	0.91	0.87	0.99
	Jogging	0.97	0.97	1.00
	Standing	0.87	0.86	0.95
	Sitting	0.89	0.98	0.88
	Biking	0.90	0.98	1.00
	Walking Upstairs	0.98	0.99	0.98
	Walking Downstairs	0.97	0.97	0.99

Table 12. *Cont.*

Type	Activity	F1-Score		
		DeepConvTCN [49]	InceptionTime [48]	The Proposed CNN-ResBiGRU-CBAM
Asymmetric	Typing	0.92	0.95	0.99
	Writing	0.98	0.91	1.00
	Drinking Coffee	0.82	0.85	0.94
	Talking	0.89	0.82	0.91
	Smoking	0.87	0.84	0.94
	Eating	0.97	0.99	0.97
	Average	0.918	0.922	0.965

5. Discussion

This part thoroughly examines the research observations discussed in Section 4.

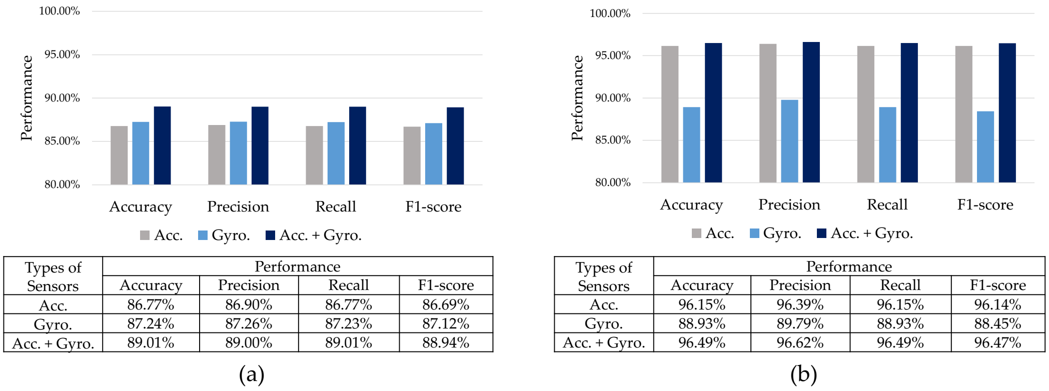
5.1. Impact of Different Types of Sensors

From reviewing the activity recognition results from the WISDM-HARB and UTwente datasets presented in Sections 3.2.1 and 3.2.2, significant insights can be derived regarding the impact of accelerometer, gyroscope, and fused sensor data on performance across various sensing modalities.

The accelerometer and gyroscope data from the WISDM-HARB dataset perform well, but the gyroscope exhibits slightly higher accuracy. This indicates the value of both data types in modeling activities, as depicted in Figure 11a. Moreover, improving sensor capabilities enhances effectiveness and validates the additional and supplementary information.

Upon analysis of the UTwente dataset, it becomes evident that accelerometer data outperform gyroscope data in activity recognition. This is indicated by the significantly higher accuracy values across all models, as illustrated in Figure 11b. However, data fusion further enhances performance, underscoring the advantages of integrating diverse measurements.

Our proposed CNN-ResBiGRU-CBAM architecture consistently outperforms all three modalities across public benchmark datasets. This highlights its effectiveness in accurately representing diverse sensor signals and applying this expertise to various datasets.



Types of Sensors	Performance			
	Accuracy	Precision	Recall	F1-score
Acc.	86.77%	86.90%	86.77%	86.69%
Gyro.	87.24%	87.26%	87.23%	87.12%
Acc. + Gyro.	89.01%	89.00%	89.01%	88.94%

(a)

Types of Sensors	Performance			
	Accuracy	Precision	Recall	F1-score
Acc.	96.15%	96.39%	96.15%	96.14%
Gyro.	88.93%	89.79%	88.93%	88.45%
Acc. + Gyro.	96.49%	96.62%	96.49%	96.47%

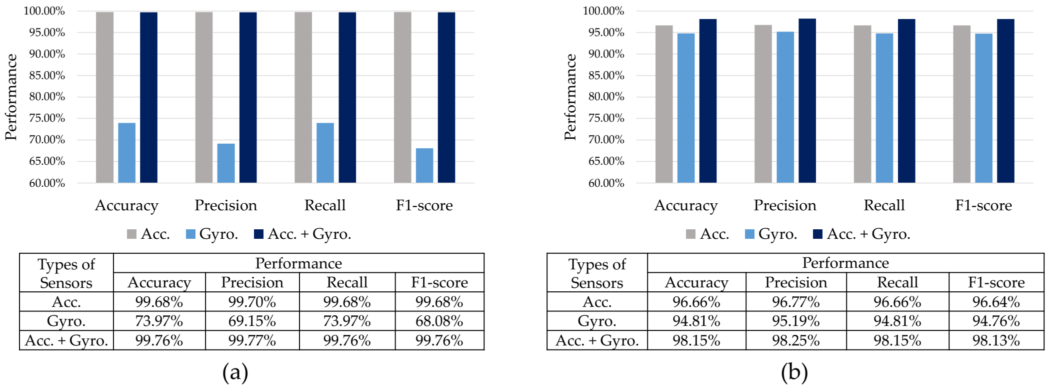
(b)

Figure 11. Comparison results of the proposed CNN-ResBiGRU-CBAM using data from different types of sensors: (**a**) WISDM-HARB dataset; (**b**) UTwente dataset.

5.2. Impact of Different Types of Activities

To assess the effectiveness of our CNN-ResBiGRU-CBAM model in identifying various activity types, we devised two distinct experimental situations. In the first scenario, the network was trained and tested using only periodic symmetric activity data, such as walking, running, and leaping. In the second scenario, only more intricate asymmetric activity data weree used, including writing, eating, and smoking.

By examining the recognition performance results shown in Figure 12, we can assess the influence of activity type, symmetric vs. asymmetric, on our CNN-ResBiGRU-CBAM model.



Types of Sensors	Performance			
	Accuracy	Precision	Recall	F1-score
Acc.	99.68%	99.70%	99.68%	99.68%
Gyro.	73.97%	69.15%	73.97%	68.08%
Acc. + Gyro.	99.76%	99.77%	99.76%	99.76%

(a)

Types of Sensors	Performance			
	Accuracy	Precision	Recall	F1-score
Acc.	96.66%	96.77%	96.66%	96.64%
Gyro.	94.81%	95.19%	94.81%	94.76%
Acc. + Gyro.	98.15%	98.25%	98.15%	98.13%

(b)

Figure 12. Comparison results of the proposed CNN-ResBiGRU-CBAM for classifying different types of activities: (**a**) symmetric activities; (**b**) asymmetric activities.

The accelerometer sensor demonstrates exceptional accuracy of 99.68% for symmetric operations, surpassing the gyroscope's accuracy of 73.97% by a significant margin. This is consistent with the idea that symmetric movements are periodic patterns that may be better understood by examining acceleration along the axes. By integrating sensors, the accuracy of symmetric activity detection is enhanced to 99.76%, demonstrating the inclusion of additional, although very insignificant, supplementary data. The model design efficiently handles these cyclic time series.

Regarding activities that are not symmetrical, the accelerometer has an accuracy of 96.66%, while the gyroscope achieves a corresponding accuracy of 94.81%. Both sensors give valuable distinguishing indications. Integrating sensor data dramatically improves the accuracy of detecting asymmetric activity, reaching a validation rate of 98.15%. This highlights the need to combine diverse data sources to identify complex contextual information accurately. Our model's precision, recall, and F1-score exhibit consistent patterns of comparison, indicating reliable and generalized performance rather than exaggerated accuracy.

6. Conclusions

In conclusion, the CNN-ResBiGRU-CBAM model proposed in this study significantly advances HAR using wearable sensors. By effectively combining residual connections, BiGRU, and channel-wise attention, our approach addresses the long-standing challenge of accurately distinguishing between symmetric and asymmetric human activities. The model's ability to automatically learn discriminative features eliminates manual feature engineering, streamlining the process and improving overall performance. By integrating a ResNet backbone and dual attention blocks, the CNN-ResBiGRU-CBAM model efficiently captures spatial feature hierarchies and prioritizes the most salient aspects of complex asymmetric activities. Furthermore, its ability to process long activity sequences overcomes the limitations of earlier recurrent models, enhancing its versatility and applicability.

Extensive evaluations on the WISDM-HARB and Utwente benchmark datasets have demonstrated the superiority of our model in accurately recognizing a wide range of activities, including periodic locomotions and intricate asymmetric gestures. Additionally, our study offers valuable insights into the patterns of confusion between symmetric and asymmetric classes and the importance of temporal attributes in activity recognition. The CNN-ResBiGRU-CBAM model sets a new standard for HAR using wearable sensors, providing a robust and precise approach for recognizing both balanced and subtle real-life human gestures. This advancement can revolutionize contextual understanding in various applications, such as wearable computing and ambient assisted living.

While the proposed model represents a significant stride forward, there remain opportunities for further research and improvement. Investigating the integration of continuous HAR and conducting additional studies into model uncertainty could enhance the robustness and reliability of the proposed strategy, paving the way for even more accurate and dependable activity recognition systems.

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

Funding: This research was funded by University of Phayao (Grant No. FF67-UoE-214); Thailand Science Research and Innovation Fund (Fundamental Fund 2024); and King Mongkut's University of Technology North Bangkok under Contract No. KMUTNB-67-KNOW-12.

Data Availability Statement: The original data presented in the study are openly available for the WISDM-HARB dataset in the UCI Machine Learning Repository at https://archive.ics.uci.edu/dataset/507/wisdm+smartphone+and+smartwatch+activity+and+biometrics+dataset (accessed on 22 Feburary 2024) and the UTwente dataset in the Pervasive Systems group at https://www.utwente.nl/en/eemcs/ps/research/dataset/(accessed on 22 Feburary 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Y.; Cang, S.; Yu, H. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **2019**, *137*, 167–190. [CrossRef]
2. Wang, Z.; Yang, Z.; Dong, T. A Review of Wearable Technologies for Elderly Care that Can Accurately Track Indoor Position, Recognize Physical Activities and Monitor Vital Signs in Real Time. *Sensors* **2017**, *17*, 341. [CrossRef] [PubMed]
3. Mostafa, H.; Kerstin, T.; Regina, S. Wearable Devices in Medical Internet of Things: Scientific Research and Commercially Available Devices. *Healthc. Inform. Res.* **2017**, *23*, 4–15. [CrossRef]
4. Ha, P.J.; Hyun, M.J.; Ju, K.H.; Hee, K.M.; Hwan, O.Y. Sedentary Lifestyle: Overview of Updated Evidence of Potential Health Risks. *Korean J. Fam. Med.* **2020**, *41*, 365–373. [CrossRef]
5. Oh, Y.; Choi, S.A.; Shin, Y.; Jeong, Y.; Lim, J.; Kim, S. Investigating Activity Recognition for Hemiparetic Stroke Patients Using Wearable Sensors: A Deep Learning Approach with Data Augmentation. *Sensors* **2024**, *24*, 210. [CrossRef]
6. Kraft, D.; Srinivasan, K.; Bieber, G. Deep Learning Based Fall Detection Algorithms for Embedded Systems, Smartwatches, and IoT Devices Using Accelerometers. *Technologies* **2020**, *8*, 72. [CrossRef]
7. Mekruksavanich, S.; Jitpattanakul, A. Deep Residual Network for Smartwatch-Based User Identification through Complex Hand Movements. *Sensors* **2022**, *22*, 3094. [CrossRef] [PubMed]
8. Proffitt, R.; Ma, M.; Skubic, M. Development and Testing of a Daily Activity Recognition System for Post-Stroke Rehabilitation. *Sensors* **2023**, *23*, 7872. [CrossRef]
9. Zhou, X.; Liang, W.; Wang, K.I.K.; Wang, H.; Yang, L.T.; Jin, Q. Deep-Learning-Enhanced Human Activity Recognition for Internet of Healthcare Things. *IEEE Internet Things J.* **2020**, *7*, 6429–6438. [CrossRef]
10. Fridriksdottir, E.; Bonomi, A.G. Accelerometer-Based Human Activity Recognition for Patient Monitoring Using a Deep Neural Network. *Sensors* **2020**, *20*, 6424. [CrossRef]
11. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209. [CrossRef]
12. Peng, L.; Chen, L.; Ye, Z.; Zhang, Y. AROMA: A Deep Multi-Task Learning Based Simple and Complex Human Activity Recognition Method Using Wearable Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 74. [CrossRef]
13. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors. *Sensors* **2016**, *16*, 426. [CrossRef] [PubMed]
14. Alo, U.R.; Nweke, H.F.; Teh, Y.W.; Murtaza, G. Smartphone Motion Sensor-Based Complex Human Activity Identification Using Deep Stacked Autoencoder Algorithm for Enhanced Smart Healthcare System. *Sensors* **2020**, *20*, 6300. [CrossRef] [PubMed]
15. Liu, L.; Peng, Y.; Liu, M.; Huang, Z. Sensor-based human activity recognition system with a multilayered model using time series shapelets. *Knowl. Based Syst.* **2015**, *90*, 138–152. [CrossRef]
16. Chen, L.; Liu, X.; Peng, L.; Wu, M. Deep learning based multimodal complex human activity recognition using wearable devices. *Appl. Intell.* **2021**, *51*, 4029–4042. [CrossRef]
17. Tahir, B.S.; Ageed, Z.S.; Hasan, S.S.; Zeebaree, S.R.M. Modified Wild Horse Optimization with Deep Learning Enabled Symmetric Human Activity Recognition Model. *Comput. Mater. Contin.* **2023**, *75*, 4009–4024. [CrossRef]

18. Cengiz, A.B.; Birant, K.U.; Cengiz, M.; Birant, D.; Baysari, K. Improving the Performance and Explainability of Indoor Human Activity Recognition in the Internet of Things Environment. *Symmetry* **2022**, *14*, 2022. [CrossRef]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
20. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842.
21. Long, J.; Sun, W.; Yang, Z.; Raymond, O.I. Asymmetric Residual Neural Network for Accurate Human Activity Recognition. *Information* **2019**, *10*, 203. [CrossRef]
22. Tuncer, T.; Ertam, F.; Dogan, S.; Aydemir, E.; Pławiak, P. Ensemble Residual Networks based Gender and Activity Recognition Method with Signals. *J. Supercomput.* **2020**, *76*, 2119–2138. [CrossRef]
23. Ronald, M.; Poulose, A.; Han, D.S. iSPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition. *IEEE Access* **2021**, *9*, 68985–69001. [CrossRef]
24. Mehmood, K.; Imran, H.A.; Latif, U. HARDenseNet: A 1D DenseNet Inspired Convolutional Neural Network for Human Activity Recognition with Inertial Sensors. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–6. [CrossRef]
25. Xu, C.; Chai, D.; He, J.; Zhang, X.; Duan, S. InnoHAR: A Deep Neural Network for Complex Human Activity Recognition. *IEEE Access* **2019**, *7*, 9893–9902. [CrossRef]
26. Zhao, Y.; Yang, R.; Chevalier, G.; Gong, M. Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors. *arXiv* **2017**, arXiv:1708.08989.
27. Malki, Z.; Atlam, E.S.; Dagnew, G.; Alzighaibi, A.; Elmarhomy, G.; Gad, I. Bidirectional Residual LSTM-based Human Activity Recognition. *Comput. Inf. Sci.* **2020**, *13*, 40. [CrossRef]
28. Challa, S.; Semwal, V. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *Vis. Comput.* **2021**, *38*, 4095–4109. [CrossRef]
29. Gao, W.; Zhang, L.; Teng, Q.; He, J.; Wu, H. DanHAR: Dual Attention Network for multimodal human activity recognition using wearable sensors. *Appl. Soft Comput.* **2021**, *111*, 107728. [CrossRef]
30. Murahari, V.S.; Plötz, T. On attention models for human activity recognition. In Proceedings of the 2018 ACM International Symposium on Wearable Computers ISWC '18, Singapore, 8–12 October 2018; pp. 100–103. [CrossRef]
31. Khan, Z.N.; Ahmad, J. Attention induced multi-head convolutional neural network for human activity recognition. *Appl. Soft Comput.* **2021**, *110*, 107671. [CrossRef]
32. Weiss, G.M.; Yoneda, K.; Hayajneh, T. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. *IEEE Access* **2019**, *7*, 133190–133202. [CrossRef]
33. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition using Smartphones. In Proceedings of the The European Symposium on Artificial Neural Networks, Bruges, Belgium, 24–26 April 2013; pp. 437–442.
34. Mekruksavanich, S.; Jitpattanakul, A. Deep Convolutional Neural Network with RNNs for Complex Activity Recognition Using Wrist-Worn Wearable Sensor Data. *Electronics* **2021**, *10*, 1685. [CrossRef]
35. Banos, O.; Galvez, J.M.; Damas, M.; Pomares, H.; Rojas, I. Window Size Impact in Human Activity Recognition. *Sensors* **2014**, *14*, 6474–6499. [CrossRef]
36. Hochreiter, S. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **1998**, *6*, 107–116. [CrossRef]
37. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014; pp. 103–111. [CrossRef]
38. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
40. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 3–19.
41. Agac, S.; Durmaz Incel, O. On the Use of a Convolutional Block Attention Module in Deep Learning-Based Human Activity Recognition with Motion Sensors. *Diagnostics* **2023**, *13*, 1861. [CrossRef] [PubMed]
42. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016.
43. Zhang, Z.; Sabuncu, M.R. Generalized cross entropy loss for training deep neural networks with noisy labels. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, Montreal, QC, Canada, 3–8 December 2018; pp. 8792–8802.
44. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
45. Wong, T.T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [CrossRef]
46. Bragança, H.; Colonna, J.G.; Oliveira, H.A.B.F.; Souto, E. How Validation Methodology Influences Human Activity Recognition Mobile Systems. *Sensors* **2022**, *22*, 2360. [CrossRef] [PubMed]

47. Suglia, V.; Palazzo, L.; Bevilacqua, V.; Passantino, A.; Pagano, G.; D'Addio, G. A Novel Framework Based on Deep Learning Architecture for Continuous Human Activity Recognition with Inertial Sensors. *Sensors* **2024**, *24*, 2199. [CrossRef]
48. Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1936–1962. [CrossRef]
49. Aparecido Garcia, F.; Mazzoni Ranieri, C.; Aparecida Francelin Romero, R. Temporal Approaches for Human Activity Recognition Using Inertial Sensors. In Proceedings of the 2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE), Rio Grande, Brazil, 23–25 October 2019; pp. 121–125. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.