applied sciences

MDPI

Article

Attention-Based Hybrid Deep Learning Network for Human Activity Recognition Using WiFi Channel State Information

Sakorn Mekruksavanich [1], Wikanda Phaphan [2], Narit Hnoohom [3,*] and Anuchit Jitpattanakul [4,5,*]

[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 Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand; wikanda.p@sci.kmutnb.ac.th
[3] Department of Computer Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom 73170, Thailand
[4] Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
[5] Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
* Correspondence: narit.hno@mahidol.ac.th (N.H.); anuchit.j@sci.kmutnb.ac.th (A.J.)

Abstract: The recognition of human movements is a crucial aspect of AI-related research fields. Although methods using vision and sensors provide more valuable data, they come at the expense of inconvenience to users and social limitations including privacy issues. WiFi-based sensing methods are increasingly being used to collect data on human activity due to their ubiquity, versatility, and high performance. Channel state information (CSI), a characteristic of WiFi signals, can be employed to identify various human activities. Traditional machine learning approaches depend on manually designed features, so recent studies propose leveraging deep learning capabilities to automatically extract features from raw CSI data. This research introduces a versatile framework for recognizing human activities by utilizing CSI data and evaluates its effectiveness on different deep learning networks. A hybrid deep learning network called CNN-GRU-AttNet is proposed to automatically extract informative spatial-temporal features from raw CSI data and efficiently classify activities. The effectiveness of a hybrid model is assessed by comparing it with five conventional deep learning models (CNN, LSTM, BiLSTM, GRU, and BiGRU) on two widely recognized benchmark datasets (CSI-HAR and StanWiFi). The experimental results demonstrate that the CNN-GRU-AttNet model surpasses previous state-of-the-art techniques, leading to an average accuracy improvement of up to 4.62%. Therefore, the proposed hybrid model is suitable for identifying human actions using CSI data.

Keywords: human activity recognition; WiFi sensing; deep learning; attention mechanism; channel state information

Citation: Mekruksavanich, S.; Phaphan, W.; Hnoohom, N.; Jitpattanakul, A. Attention-Based Hybrid Deep Learning Network for Human Activity Recognition Using WiFi Channel State Information. *Appl. Sci.* **2023**, *13*, 8884. https://doi.org/10.3390/app13158884

Academic Editors: Luigi Bibbò and Marley M.B.R. Vellasco

Received: 29 May 2023
Revised: 27 July 2023
Accepted: 29 July 2023
Published: 1 August 2023

1. Introduction

Over the last decade, human activity recognition (HAR) research has advanced significantly. It has proven successful in several areas, such as healthcare, smart homes, sports performance tracking, and human–computer interaction [1]. The primary goal of HAR is to detect and understand user actions, enabling computing systems to provide proactive assistance [2,3]. There are two main categories of HAR: vision-based and sensor-based. Firstly, vision-based HAR (V-HAR) holds great promise, benefiting from the rapid advancements in computer vision techniques and the high resolution offered by optical sensors [4–6]. Despite its success, V-HAR still faces challenges such as illumination, occlusion, and privacy concerns. However, sensor-based HAR (S-HAR) has become increasingly popular due to the rapid advancement in sensor technology [7–9]. S-HAR collects data from low-level

sensors, such as accelerometers, gyroscopes, magnetometers, and acoustic sensors, to extract high-level information about human behavior. However, S-HAR has limitations in terms of environmental requirements, and people may object to using sensors due to their bothersome or cumbersome nature. V-HAR and S-HAR have challenges to overcome, but they can potentially provide valuable insights into human behavior.

Despite the numerous methods developed in recent years, WiFi-based sensing techniques have gained significant attention due to their widespread availability, versatility, and high performance [10]. WiFi-based sensing has the potential to integrate sensing and communication functions, as channel information can be utilized for both purposes [11]. Compared to V-HAR and S-HAR techniques, WiFi-based HAR systems provide several advantages. WiFi-based HAR systems differ from V-HAR systems in that they are not influenced by lighting conditions or variations in human body shapes, and they also respect user privacy. Additionally, these systems provide a more convenient option for smart home and healthcare applications since they do not rely on users wearing sensors. Consequently, researchers have actively engaged in investigating and developing WiFi-based HAR methods in recent times.

Wi-Fi-based human HAR systems offer a cost-effective and seamless integration solution within existing Wi-Fi infrastructures in both residential and commercial environments, with minimal additional expenses. These systems can be arranged into two main types [12] based on their utilization of the received signal strength indicator (RSSI) [13], while the other type relies on channel state information (CSI) [10,14] for activity recognition tasks. The CSI provides a comprehensive characterization of the radio frequency (RF) signal propagation, encompassing aspects such as amplitude attenuation, time lag, and phase shift across various carrier frequencies. Prior research has consistently demonstrated the superior performance of CSI-based HAR systems compared to RSSI-based alternatives [14], primarily due to the increased richness and informational content provided by CSI data.

Learning-based approaches have emerged as potent tools for classification and prediction, occupying a crucial role in HAR and the implementation of recognition models. Researchers have extensively employed conventional machine learning (ML) techniques, including Hidden Markov Model [15], Random Forest [16], Support Vector Machine [17], and K-Nearest Neighbor [18], to achieve HAR objectives. In conventional activity recognition methods, ML algorithms manually extract features from sensor data, often relying on statistical or structural attributes such as means, medians, and standard deviations. Extracting the most relevant manual features often demands domain expertise. While these handcrafted features demonstrate satisfactory performance in scenarios with limited training data, their extraction becomes increasingly intricate as the number of sensors escalates.

Deep learning (DL), a cutting-edge approach within the realm of ML, has gained significant traction due to its remarkable capability to extract features and perform classification simultaneously. In contrast to traditional ML methods, DL leverages artificial neural networks with multiple layers to process data and address intricate problems. Promising outcomes have been observed across various domains through DL models including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNN). CNN effectively overcomes high dimensionality by employing filters capable of convolving and sharing weights. On the other hand, RNN, a type of neural network, leverages previous outputs as inputs in the present step and incorporates hidden states, making it particularly suitable for solving sequential concerns and operating time series data. DL presents distinct advantages over traditional ML approaches as it surpasses the limitations of manual feature extraction and exhibits enhanced efficiency in handling large datasets. Additionally, Graphics Processing Units (GPUs) can be employed to accelerate the computational speed of DL models.

Over the past few years, the improvement of DL techniques has become increasingly prominent in CSI-based HAR [19]. Among these techniques, the LSTM method has emerged as a notable strategy that relies solely on the current state of CSI for learning. LSTM excels at autonomously learning representative features and capturing temporal data

during the feature learning process. To attain superior performance in HAR utilizing CSI measurements, researchers have developed an attention-based bidirectional LSTM method. This method combines a Bidirectional LSTM (BLSTM) [20] with an attention model that assigns increased weights to specific time steps, effectively enhancing recognition efficiency.

The conventional LSTM is limited to processing sequential CSI measurements in a single direction, such as forward, and it solely relies on historical CSI information for the current hidden state. However, we emphasize the significance of incorporating future CSI data to enable accurate HAR. Additionally, the sequentially learned features of a conventional LSTM can have varying effects on the HAR task. In the conventional LSTM method, each learned characteristic contributes equally to the final identification of human activities. Real-time applications can benefit from employing advanced DL approaches and models to enhance the accuracy of these methods.

Therefore, this article presents CNN-GRU-AttNet, an innovative DL network specifically designed for extracting spatial-temporal features from raw WiFi CSI data. The network architecture comprises convolution layers and a gated recurrent unit (GRU) layer. Moreover, we incorporate an attention mechanism that dynamically assigns weights to important features and time steps, thereby enhancing the model's generalization performance for HAR. To evaluate the effectiveness of our proposed model, we conduct a comprehensive set of experiments and compare its performance against existing benchmark approaches. The main contributions of this research can be succinctly summarized as:

- Development of a novel DL framework that enables HAR using WiFi CSI measurements, eliminating the need for manual feature extraction.
- Introduction of a hybrid DL network, CNN-GRU-AttNet, that leverages the strengths of CNN and GRU to automatically extract spatial and temporal features, leading to highly accurate HAR results.
- Integration of an attention mechanism into the CNN-GRU-AttNet network, allowing for the prioritization of important features and time steps, thereby enhancing recognition performance.
- Thorough evaluation of the proposed approach through a series of rigorous experiments, demonstrating its superior performance in HAR using WiFi CSI data.

The paper follows the subsequent structure: Section 2 presents an extensive review of existing research on HAR utilizing WiFi CSI data. In Section 3, we introduce the framework for automatic learning and selection of features in the HAR process, along with the detailed description of the proposed CNN-GRU-AttNet model. The experimental setup and the results obtained under various scenarios are outlined in Section 4. Section 5 provides an in-depth discussion of the experimental findings, analyzing their implications and significance. Lastly, Section 6 concludes the study by summarizing the key findings and suggesting possible avenues for planned research endeavors.

2. Related Works

2.1. CSI-Based HAR

Within the existing literature, numerous WiFi-based HAR systems have been researched and analyzed, capitalizing on the widespread availability of WiFi signals. Notably, Abdelnasser et al. [21] presented a system called WiGest, which consists of three integral components: initial feature extraction, gesture recognition, and motion mapping. This system employs RSS measurements for accurate gesture identification. Additionally, Gu et al. [22] proposed an alternative approach that leverages WiFi RSS to recognize human activities. Through manual extraction of significant features from raw RSS measurements, they introduced a fusion algorithm capable of identifying essential movements such as standing and walking.

The efficacy of activity recognition mechanisms based on RSS measurements is limited by the presence of instability and disorder caused by multi-path and fading effects, even when considering basic activities. While RSS provides a broad understanding of communication links, CSI offers more intricate details about the condition of the communication

channel [23]. Notably, the enhanced reliability and informativeness of WiFi CSI have attracted considerable attention. Zhang et al. [24] devised a Fresnel zone model for HAR that employs WiFi CSI signals, enabling the assessment of WiFi signals' sensing capabilities. Through their proposed model, they achieved remarkable accuracy in detecting human behaviors at centimeter and decimeter scales, such as respiration rate and the orientation of walking. Furthermore, Wang et al. [25] introduced a location-based movement identification system that utilizes WiFi CSI readings.

Previous studies relied on hand-crafted features, which obligate expert knowledge and may not capture the implicit features necessary for accurate HAR using WiFi CSI. To address this issue, some researchers have proposed using DL techniques to automatically learn essential features for this task.

2.2. DL for HAR

DL techniques have gained significant traction in utilizing WiFi CSI for the purposes of localizing and classifying human actions, leveraging the wealth of wireless link information it offers. Wang et al. [26] presented an indoor localization of the HAR system based on a multitasking 1D-CNN architecture enhanced with residual connections. Their model achieved a notable accuracy of 95.68% when tested on a dataset comprising six distinct categories of human behavior. Moshiri et al. [27] gathered CSI data from various human activities and converted them into RGB images, which were then passed through a 2D-CNN layer for classification. Their best-performing model obtained an accuracy of 95%. Chahoushi et al. [28] presented a MIMO-AE for physical activity classification, which achieved a high accuracy of 94.49% using only 50% of the training data.

In HAR, RNNs and their subsets, such as LSTM, have been commonly used for CSI data analysis. However, when analyzing long sequences, these networks face problems, leading to vanishing gradients. Even with the inclusion of long memory and switch gates in LSTM, the problem persists [20]. The memory bandwidth requirements of LSTMs are substantial by reason of the complexity of their sequential direction and MLP layers. Furthermore, these models encounter difficulties when confronted with sequences comprising a large number of terms, as their performance becomes compromised beyond 100 terms [29]. Additionally, LSTMs are restricted to analyzing sequential data in a single direction, limiting their ability to capture bidirectional dependencies. Therefore, they cannot differentiate between activities such as lying and sitting down.

To overcome these limitations, researchers have developed new methods for HAR. Yousefi et al. [19] created the StanWiFi dataset, extracted statistical features, and used hidden Markov models, LSTM, and RF models to classify activities with reported accuracies of 64.6%, 73.3%, and 90.5%, respectively. The BiLSTM architecture was meticulously designed to leverage both historical and prospective CSI data [30], facilitating effective feature learning in the realm of classification. Additionally, the ABLSTM algorithm [20] underwent rigorous evaluation and comparative analysis against alternative algorithms. Zhang et al. [31] introduced the Dense-LSTM method, which demonstrated a remarkable accuracy of approximately 90% while employing a reasonable amount of CSI data. Shang et al. [32] proposed a DL model that combined LSTM-CNN with WiFi CSI signals, yielding an average performance of 94.14% on a publicly available dataset. Moreover, Santosh et al. [33] presented CSITime, an adjusted InceptionTime structure customized for HAR tasks using WiFi CSI signals, achieving an impressive accuracy of 98% on the StanWiFi dataset.

As discussed earlier, several researchers have explored different techniques for HAR, including feature extraction, CNN, and RNN-based models, as shown in Table 1. The presence of spatial and temporal characteristics within WiFi CSI-based HAR data necessitates the utilization of a model capable of effectively capturing both aspects of human behavior. To address this, our study introduces a hybrid DL model that is proficient in learning spatial-temporal features, all while maintaining a streamlined parameter count. This model offers enhanced precision and accuracy for HAR tasks.

Table 1. Previous works analysis using DL approaches for HAR based on CSI data.

Year	Classifier Model	Dataset	Physical Activities	Accuracy
2019	1D-CNN [26]	Private	six casual activities	88.13%
2017	LSTM [19]	StanWiFi	lie down; fall; walk; run; sit down; stand up	90.50%
2019	ABLSTM [20]	StanWiFi	lie down; fall; walk; run; sit down; stand up	97.30%
2020	Dense-LSTM [31]	Private	empty; jump; pick up; run; sit down; wave hand; walk make phone call; check wristwatch; walk normal and fast; run; jump; lie down; play guitar and piano; play basketball	90.00%
2021	LSTM-CNN [32]	Private	stand; sit, falling; standing up; stepping	94.14%
2021	2D-CNN [27]	CSI-HAR	lie down; fall; bend; run; sit down; stand up; walk	95.00%
2022	CSITime [33]	StanWiFi	lie down; fall; walk; run; sit down; stand up	98.00%
2023	MIMI-AE [28]	CSI-HAR	lie down; fall; bend; run; sit down; stand up; walk	94.49%

3. Proposed Methodology

This research introduces an HAR system that utilizes a hybrid DL network called CNN-GRU-AttNet based on WiFi CSI data. The first step involves collecting raw CSI data for DL networks. The raw CSI data are pre-processed in the second step using denoising and segmentation techniques. Following that, the pre-processed CSI data are partitioned into separate training and evaluating sets utilizing a five-fold cross-validation methodology. Subsequently, the data samples undergo a process of high-dimensional embedding to generate features by employing convolutional layers and a GRU layer within the CNN-GRU-AttNet model. Finally, the system's performance is evaluated using standard assessment techniques, such as accuracy, precision, recall, and F1-score. Figure 1 illustrates the overall organization of the framework.

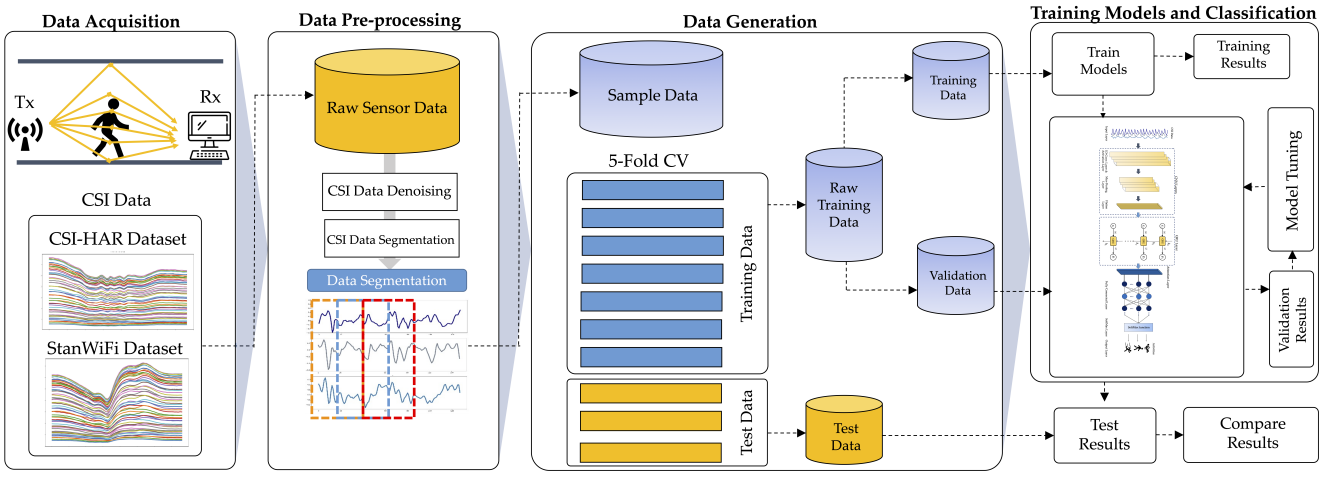

Figure 1. A CSI-based HAR framework using a hybrid DL network.

3.1. Data Acquisition

This study conducted experiments using two publicly available datasets: CSI-HAR and StanWiFi. The details of both datasets are presented in Table 2.

Table 2. Summary of the CSI datasets used in this study.

Dataset	No. of Participants (Age Range)	Collection Tools	Bandwidth and Number of Subcarries	Activities	No. of Samples
CSI-HAR	3 (25 to 70 yrs)	Raspberry Pi-4B Nexmon CSI Tool	40 MHz and 52 Subcarriers	Lie down	405
				Fall	437
				Bend	415
				Run	449
				Sit down	413
				Stand up	348
				Walk	398
StanWiFi	6 (unidentified)	Intel 5300 NIC	20 MHz and 30 Subcarriers	Lie down	657
				Fall	443
				Run	1209
				Sit down	400
				Stand up	304
				Walk	1465

3.1.1. CSI-HAR Dataset

The proposed model's performance and comparable baseline models for WiFi-based HAR were evaluated using the publicly available CSI-HAR dataset [27]. The dataset was collected by building in the Nexmon tool on a Raspberry Pi-4GB, which allowed for collecting and storing CSI data based on transmitted and received information. The dataset contains 4000 CSI samples collected over 20 s, with each line representing 5 ms. The activity-related parts of the data were separated and stored in CSV files as matrices with 52 columns and 600 to 1100 rows, depending on the activity time. Along with the CSI samples, label files were provided to distinguish the lines for each action. The dataset consists of seven discrete activities, namely walk, run, sit down, lie down, stand up, bend, and fall. These actions were operated a total of twenty repetitions by three participants across different age groups within a controlled homeroom environment.

3.1.2. StanWiFi Dataset

Within the StanWiFi dataset [19], there are seven distinct activities: lie down, fall, walk, run, sit down, stand up, and pick up. These activities were performed by 6 participants, and each activity was repeated 20 times. The data gathering involved a Wi-Fi router with a single antenna transmitting signals, while a laptop equipped with NIC-5300 Intel's network interface card and three antennas received the signals. The transmitter and receiver were positioned 3 m away from each other in a line-of-sight scenario, and the duration of each activity was set at 20 s. With a sampling frequency of 1000 Hz, the dataset incorporated an input feature vector that encompassed both raw CSI amplitude data and a 90-dimensional vector. This vector consisted of 3 antennas and 30 subcarriers. The original dataset had seven categories, but only six were used in this study to facilitate comparison with previous works. The majority of methods proposed in the literature (e.g., [19,20,33]) have been evaluated on six activity classes from the dataset, with the "pick up" activity class being excluded. In our case, to ensure fair comparison, we evaluate our proposed CNN-GRU-AttNet on the same six daily activity classes of the StanWiFi dataset. A single training datapoint is the number of samples $\times$ the number of features (500) $\times$ the number of timestamps (90).

3.2. Data Pre-Processing

3.2.1. Data Denoising

To effectively address the impact of noise on the CSI and overcome the potential lack of discernible characteristics for different activities, it is crucial to employ ML techniques for noise filtering and feature extraction. Various noise reduction techniques can be utilized, such as the implementation of Butterworth low-pass filters [34]. Nonetheless, the presence

of high-bandwidth burst and impulse noises in the CSI, using low-pass filters alone, is not feasible for achieving a seamless CSI stream.

Based on empirical evidence, there are more effective approaches to achieve this objective, including employing principal component analysis (PCA) for noise denoising [34]. PCA is a method that reduces the complexity of a system by identifying key features where a significant portion of relevant information is concentrated. In the context of PCA-based denoising, this study adheres to the recommendation proposed in [34]. It involves excluding the initial principal component and instead selecting the subsequent five principal components for feature extraction. The reason behind this choice is that the noises arising from internal state transitions are present in all CSI streams. These noises, which show a strong correlation, get mixed into the initial principal component, along with the signal generated by human motion. However, it is important to note that all the human motion signal data present in the initial principal component are also captured within the remaining principal components. In the context of PCA, the components derived from PCA show no correlation with each other. As a result, the initial principal component solely represents one of these orthogonal components, while the rest are preserved within the subsequent PCA components. Therefore, removing the initial principal component does not compromise any relevant data. The decision to choose five principal components for feature extraction is supported by empirical evidence. The aim is to find a balance between classification effectiveness and computational overhead.

To mitigate noise, the initial principal component is excluded, and the subsequent five components are utilized for feature extraction. This approach preserves data related to the dynamic reflection of the mobile target, as it is also captured in other primary components. Following the application of PCA denoising to the CSI data, specific characteristics are extracted to enhance its usability for classification purposes. To demonstrate the denoising performance, we compared the PCA denoising method using the Signal-to-Noise Ratio (SNR), which represents the ratio of signal power (meaningful information) to noise power. The denoising results are presented in Figures 2 and 3.

Figure 2 illustrates the amplitude received from all subcarriers of CSI data obtained from the CSI-HAR dataset after noise reduction using the PCA denoising method. Notably, the denoised CSI signals exhibit higher SNR values compared to the raw CSI samples. These findings indicate the successful reduction of noise from the raw CSI data.

Figure 3 presents the raw and smoothed CSI signals for six human activities from the StanWiFi dataset. The visualizations reveal similar SNR results across all subcarriers of the CSI data after noise reduction using the PCA denoising method. Similar to the previous investigation, the SNR value of the denoised CSI data is higher than the SNR value of the raw data.

3.2.2. Segmentation

Segmentation is a crucial process that involves dividing a signal into smaller sections or windows. In our research, we utilize segmentation for two primary purposes. The first challenge we encounter is the variability in the captured CSI signals, which can differ in length and belong to different subjects. This variability hinders the identification procedure. The second challenge relates to the temporal aspect of processing an extensive volume of CSI data, which requires significant time and computational resources.

To address these challenges effectively, our study adopts a predetermined window size. This window size allows us to partition the denoised CSI signal into multiple smaller signals. By doing so, we can treat each small signal as an independent instance during the training phase of the CNN-GRU-AttNet model. This approach not only enhances efficiency but also improves the accuracy of our results.

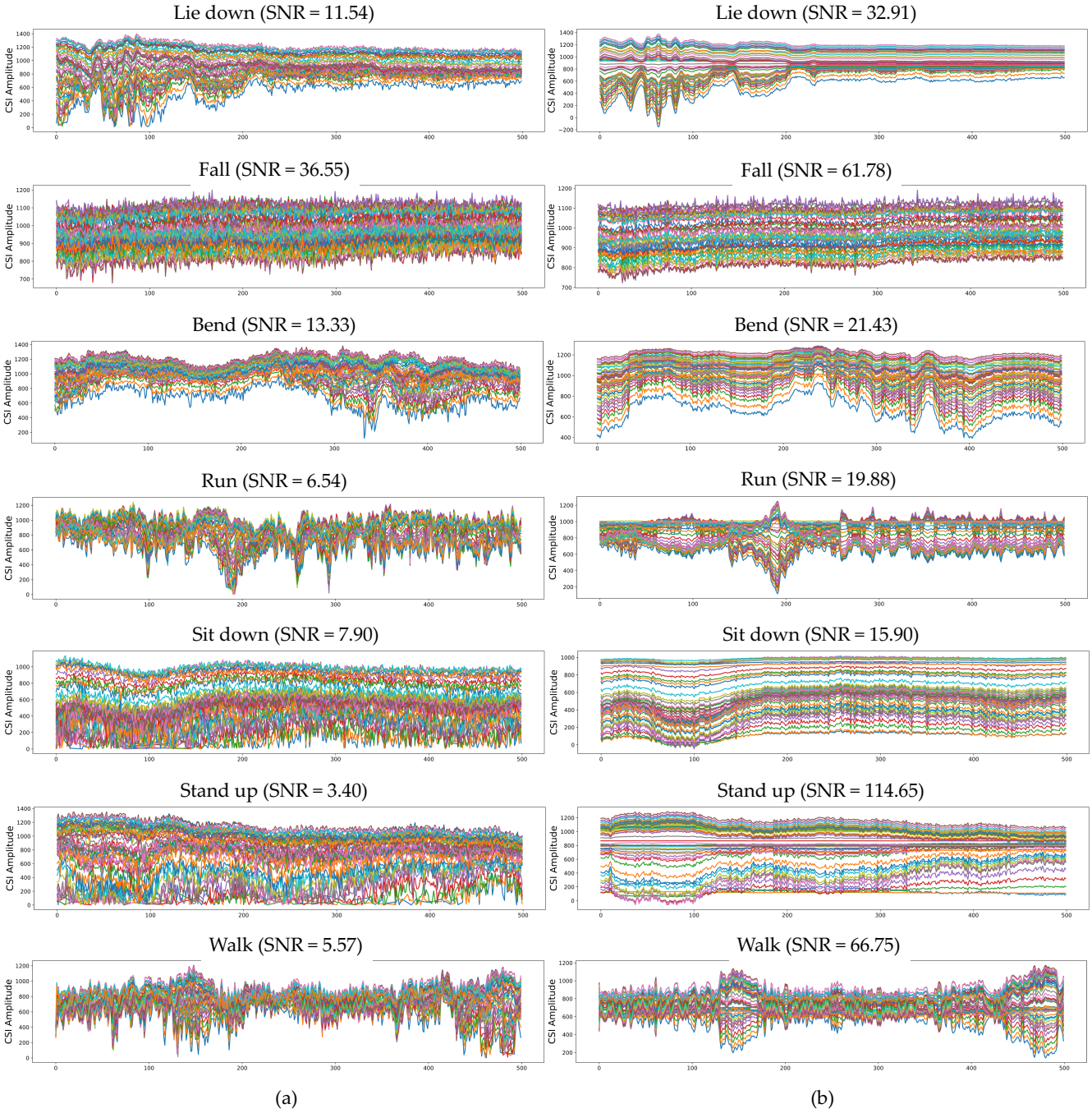

Figure 2. Some CSI signal representations from CSI-HAR dataset after pre-processing: (**a**) before denoising; (**b**) after denoising.

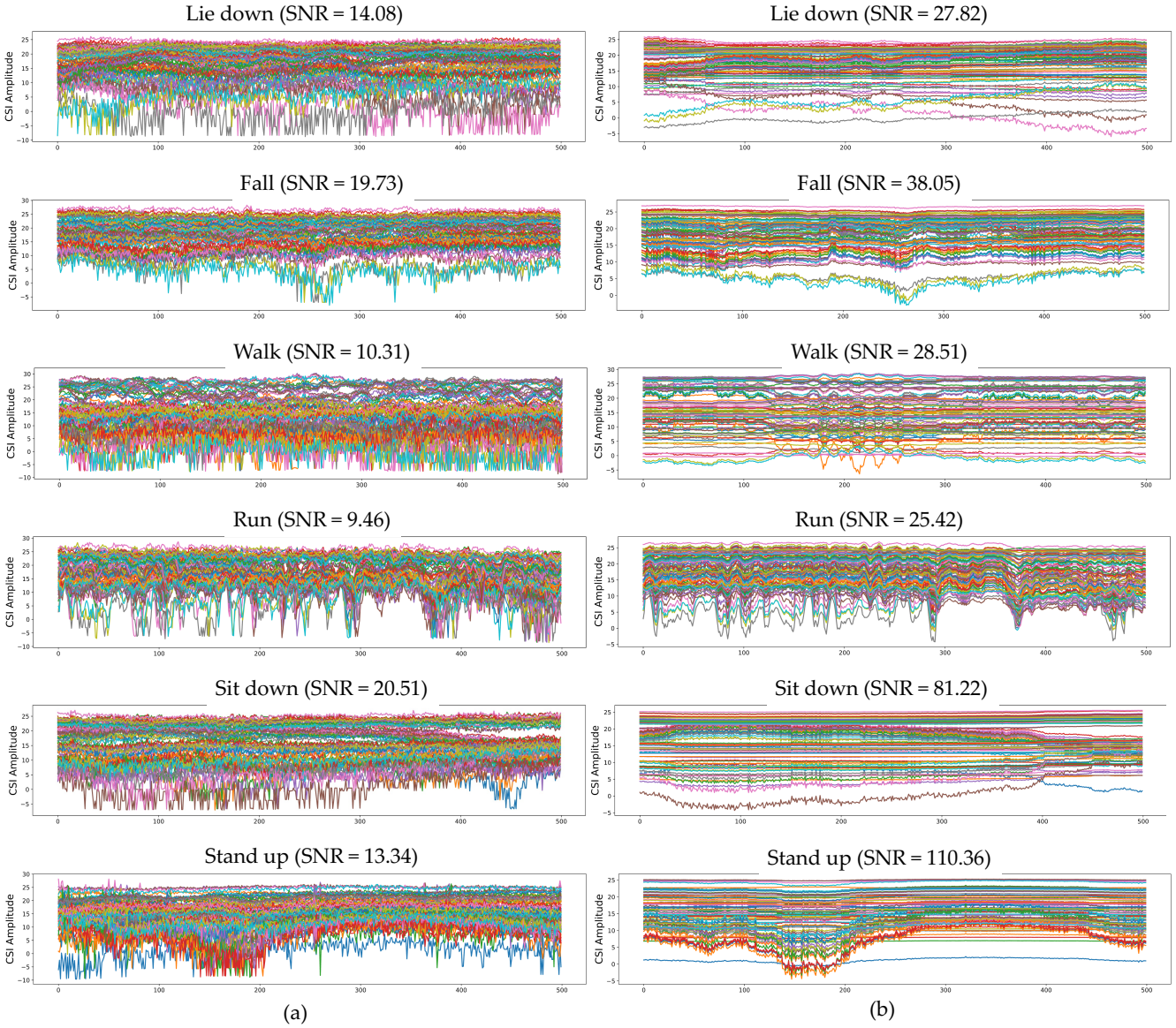

Figure 3. Some CSI signal representations from StanWiFi dataset after pre-processing: (**a**) before denoising; (**b**) after denoising.

3.3. Recognition Model

This section presents CNN-GRU-AttNet, an attention-based neural network designed for recognizing human activities using WiFi CSI data, as illustrated in Figure 4. The proposed CNN-GRU-AttNet comprises five layers: the input layer, two CNN layers, a GRU layer, an attention layer, a fully connected layer, and an output layer. Each of these layers will be described in detail below.

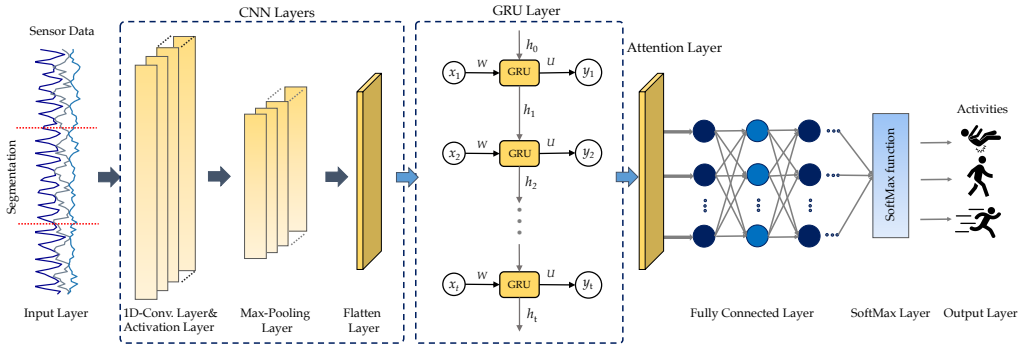

Figure 4. The proposed CNN-GRU-AttNet architecture for CSI-based HAR in this work.

CNNs extensively employ DL models with robust feature extraction capabilities. They can effectively and automatically extract features from input data, especially two-dimensional image data, and process them quickly. The convolutional layers in CNN are different from traditional neural network models since they are not fully connected. Instead, the inputs are linked to the following layers, and subregions in the input sets have the exact weights, resulting in spatially related outputs. In contrast, traditional neural network models have different weights for each input, increasing the input dimensionality and making the network more intricate. CNN addresses this issue by reducing the number of connections and weights through weight sharing and downsampling operations.

This study utilized a CNN with two layers. The first layer contained 64 filters and a kernel size of 3, while the second layer had 64 filters and a kernel size of 5. The max-pooling layers had a uniform pool size of 2. To connect the CNN and GRU layers, a flattened layer was inserted. Table 3 displays the detailed parameters of the CNN used in this research.

Table 3. Parameters of each layer of the CNN network.

Layer Name	Kernel Size	Kernel Number	Padding	Stride
Conv1D-1	5	64	2	4
Maxpooling-1	2	None	0	1
Conv1D-2	7	64	2	1
Maxpooling-2	2	None	0	1

While CNNs have proven to be highly effective in feature extraction, their performance in tasks involving time-dependent inputs, such as the analysis of biometric signal data in this study, may be relatively limited. In scenarios where sequential data are processed, the network's prediction of future states is influenced by the previous state of the input. Therefore, the network needs to consider both the current and preceding inputs. To address this challenge, the RNN model is capable of analyzing each element of the temporal sequence and incorporating both the current and preceding inputs for the current input of the RNN. The output of an RNN at a specific time step t depends on the output of the RNN at the previous time step, $t-1$.

Theoretically, RNNs are capable of acquiring knowledge from time series data with arbitrary lengths. However, when dealing with extensive time series in real-world applications, RNNs encounter the problem of gradient disappearance, which impedes the learning of long-term dependencies. To tackle this issue, we integrated a GRU as the memory component within the RNN architecture. The organization of the GRU cell's internal structure is visualized in Figure 5.

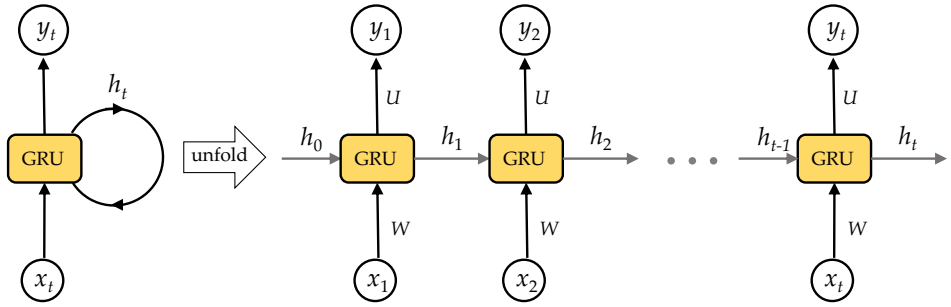

Figure 5. The structure of a GRU network.

GRU networks can be considered as a simplified form of LSTM networks within the class of RNNs, as illustrated in Figure 6. They offer enhanced computational efficiency while preserving the effectiveness of LSTM networks.

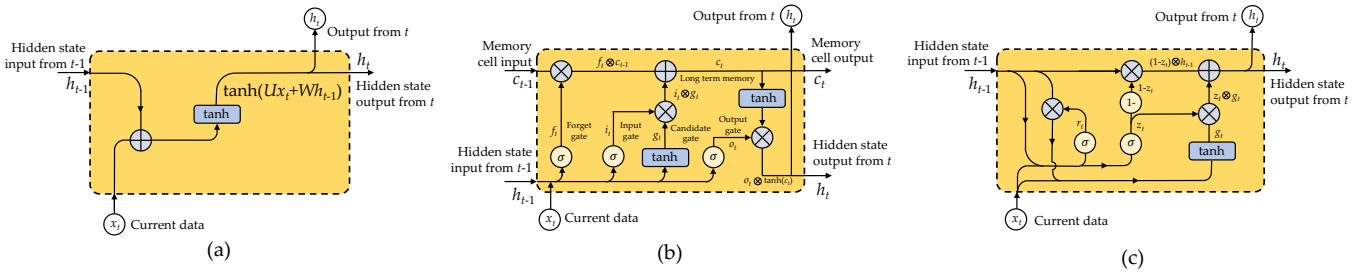

Figure 6. Comparison of RNN-based models: (**a**) simple RNN, (**b**) LSTM, and (**c**) GRU.

The architectural representation of a GRU unit, as shown in Figure 6c, consists of an update gate and a reset gate that control the extent of modification for each hidden state. These gates serve as mechanisms to regulate the flow of relevant and irrelevant information between consecutive states in a computational model. Computation of the hidden state h_t at a specific time t incorporates the update gate output z_t, the reset gate output r_t, and the current input x_t. Additionally, the preceding hidden state h_{t-1} is taken into account, as demonstrated below:

$$z_t = \sigma(W_z x_t \oplus U_z H_{t-1}) \tag{1}$$

$$r_t = \sigma(W_r x_t \oplus U_r H_{t-1}) \tag{2}$$

$$g_t = \tanh(W_g x_t \oplus U_g(r_t \otimes h_{t-1})) \tag{3}$$

$$h_t = ((1 - z_t) \otimes h_{t-1}) \oplus (z_t \otimes g_t) \tag{4}$$

The symbol σ denotes the sigmoid function, $\oplus$ denotes the elementwise addition operation, and $\otimes$ denotes the elementwise multiplication operation.

Once the GRU network has captured the contextual features, this study proposes using a self-attention mechanism to capture crucial information further. This mechanism assigns more weight to important information, leading to a more precise understanding of sequence semantics. The calculation process for the self-attention mechanism is depicted in Figure 7.

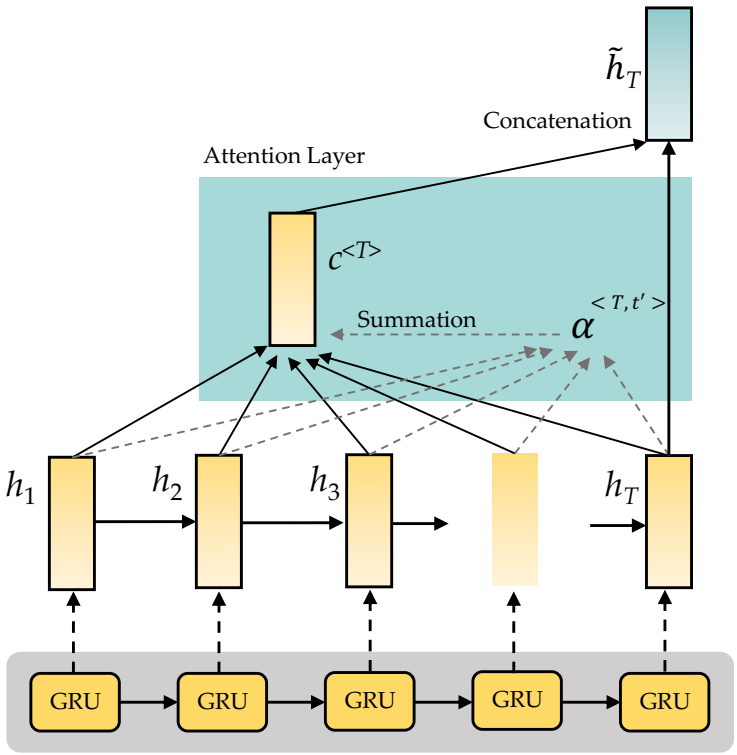

Figure 7. Attention-based GRU for the classification process.

Once the GRU layer has computed the pre-processed data $X = (x_1, x_2, ..., x_T)$, we can derive the vector $H = [h_1, h_2, h_3, ..., h_t, ..., h_T]$, where T denotes the length of the vector data X, and h_t denotes the hidden state of the GRU at time step t. We can build the self-attention mechanism for the GRU using the following steps:

$$\gamma_t = \tanh(w_2 h_t + b_2) \tag{5}$$

$$\beta_t = \frac{\exp((\gamma_t)^T w_2)}{\sum_t \exp((\gamma_t)^T w_2)} \tag{6}$$

$$\delta = \sum_t \beta_t h_t, \tag{7}$$

where w_2 is a contextual vector at the time level, β_t is a weight normalized through a softmax function, and δ represents the uniform representation of the entire sequence, which is calculated by summing all the hidden states weighted by their corresponding attention weights.

Following the attention layer, the neural network incorporates three dense layers with dropout regularization. The initial layer consists of 128 neurons and utilizes a dropout rate of 0.25. This is followed by a layer of 64 neurons with a dropout rate of 0.25 as well. Finally, the output layer of the model consists of two neurons. The rectified linear unit (ReLU) activation function is employed in all layers of the model. To achieve the best results during the training process, a configuration of 200 epochs and a batch size of 32 were utilized. The categorical cross-entropy loss function was used, and optimization was performed using the Adam optimizer [35].

3.4. Hyperparameter and Training

A three-step process is involved in building any statistical classification model. Firstly, the model development phase involves choosing hyperparameters such as batch size, activation function, learning rate, number of iterations, etc. that influence how well the model is built and trained. Adequate variation and a sufficient quantity of data are necessary for this phase. Secondly, model training and validation are carried out, with the training set being used to select hyperparameters, and the validation set to evaluate performance. In this particular instance, the training hyperparameters were carefully chosen. They consisted of a learning rate of 1×10^{-3}, 100 epochs, and a batch size of 128. To ensure efficient learning, a callback monitor was utilized to adjust the learning rate, reducing it by 75% if no progress was made for ten successive epochs. The training process incorporated data shuffling by randomizing the order of the data before the beginning of each epoch, introducing diversity. The hyperparameters were determined through an iterative process of experimentation and refinement, aiming to achieve the highest level of accuracy.

In order to assess the efficacy of the proposed model, we utilized two publicly accessible datasets, CSI-HAR and StanWiFi. Since these datasets did not have predefined training and testing sets, we adopted the five-fold cross-validation technique [36] to assess the model's performance. This technique involved randomly dividing the complete dataset into ten equally sized subsets that were mutually exclusive. The model fitting process followed an iterative procedure, where nine subsets were used for training, while the remaining subset was used for evaluating the performance. This testing and training process was repeated ten times to ensure that each subset underwent a precise testing phase. The overall performance of the model was evaluated by computing the mean value of the outcomes obtained from all iterations.

The Adam optimizer [35] played a crucial role in our methodology by updating the weights of our model. Moreover, we employed the cross-entropy loss function to quantify the error or loss during the training phase.

3.5. Network Training and Evaluation Metrics

A valuable tool for assessing the recognition performance of DL models is the confusion matrix, which provides a clear and visual representation of their performance. The multi-class confusion matrix can be mathematically represented: the rows represent the instances in the predicted class, while the columns represent the instances in the actual class.

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \dots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{bmatrix}$$

The confusion elements for each class are given by:

- True positive: $TP(C_i) = C_{ii}$;
- False positive: $FP(C_i) = \sum_{l=1}^{n} c_{li} - TP(C_i)$;
- False negative: $FN(C_i) = \sum_{l=1}^{n} c_{il} - TP(C_i)$;
- True negative: $TN(C_i) = \sum_{l=1}^{n} \sum_{k=1}^{n} c_{lk} - TP(C_i) - FP(C_i) - FN(C_i)$.

The evaluation of the DL models utilized in this study involved analyzing a confusion matrix and calculating four commonly used metrics: accuracy, precision, recall, and F1-score.

Accuracy is a measure of systematic error and is calculated by dividing the sum of true positive and true negative by the total number of records. Precision is determined by computing the ratio of examples that are correctly classified as belonging to a specific smartwatch user's class to all examples that are classified as belonging to that class. Recall is evaluated as the ratio of examples that are classified as belonging to a specific smartwatch

user's class to all examples that actually belong to that class. Lastly, the F1-score is a metric that blends precision and recall using the harmonic mean.

The mathematical expressions for these evaluation metrics were written as:

$$Accuracy = \frac{1}{|Class|} \times \sum_{i=1}^{|Class|} \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i} \tag{8}$$

$$Precision = \frac{1}{|Class|} \times \sum_{i=1}^{|Class|} \frac{TP_i}{TP_i + FP_i} \tag{9}$$

$$Recall = \frac{1}{|Class|} \times \sum_{i=1}^{|Class|} \frac{TP_i}{TP_i + FN_i} \tag{10}$$

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

4. Experiments and Findings

In this section, we provide a detailed analysis of the experiments conducted and the results obtained using the CNN-GRU-AttNet model on two distinct datasets: CSI-HAR and StanWiFi. Our aim is to demonstrate the effectiveness of the proposed model. Furthermore, we perform a comparative evaluation by assessing the performance of five baseline deep learning models (CNN, LSTM, BiLSTM, GRU, and BiGRU), along with other contemporary models, on the same datasets. This comparative analysis allows us to gain insights into the relative strengths and weaknesses of different models in the context of the given datasets.

4.1. Experimental Setting

The deep learning networks employed in this study were exclusively developed and trained on the Google Colab Pro+ platform. To expedite the model training procedure, we utilized the Tesla V100-SXM2-16GB graphics processor component. The proposed model and the standard deep learning models were implemented using the Python programming language, with Tensorflow and CUDA backend frameworks serving as the backbone. Throughout the investigation, we focused on the following Python libraries:

- To facilitate the comprehension, manipulation, and analysis of sensor data, we employed Numpy and Pandas for efficient data manipulation.
- For effective presentation and visualization of data exploration and model evaluation results, we utilized Matplotlib and Seaborn.
- In our experimental procedures, we leveraged the Scikit-learn library as a tool for data sampling and generation.
- The instantiation and training of the DL models were carried out utilized the TensorFlow, Keras, and TensorBoard frameworks.

4.2. Experimental Findings on CSI-HAR Dataset

The results of the CSI-based HAR dataset demonstrate the superior classification capabilities of the CNN-GRU-AttNet model, as presented in Table 4. The findings highlight the CNN-GRU-AttNet model's outstanding performance, with an average accuracy of 99.62%, precision of 99.61%, recall of 99.61%, and F1-score of 99.61% across all human movements. Furthermore, a comparative analysis indicates that the CNN-GRU-AttNet model exhibits exceptional efficacy in classifying HAR tasks, surpassing the achievement of the five baseline DL models.

Table 4. Performance results of both the proposed CNN-GRU-AttNet model and the five baseline models on the CSI-HAR dataset.

Model	Recognition Performance (Mean ± Std)			
	Accuracy	**Precision**	**Recall**	**F1-Score**
CNN	95.67% (±1.57%)	95.97% (±1.46%)	95.76% (±1.51%)	95.66% (±1.60%)
LSTM	84.12% (±1.22%)	84.56% (±1.14%)	84.17% (±1.24%)	84.10% (±1.18%)
BiLSTM	90.44% (±0.86%)	90.49% (±0.86%)	90.38% (±0.84%)	90.34% (±0.88%)
GRU	89.21% (±2.86%)	89.14% (±2.92%)	89.12% (±2.87%)	89.06% (±2.89%)
BiGRU	95.39% (±0.92%)	95.38% (±0.94%)	95.37% (±0.96%)	95.31% (±0.95%)
CNN-GRU-AttNet	99.62% (±0.26%)	99.61% (±0.26%)	99.61% (±0.27%)	99.61% (±0.26%)

Figure 8 illustrates the confusion matrix of the CSI-HAR dataset based on the proposed CNN-GRU-AttNet model. The matrix's diagonal elements correspond to the model's accuracy in identifying individual human actions. The findings indicate that the CNN-GRU-AttNet model is effective in capturing both the spatial and temporal features of the WiFi CSI signal. Specifically, the model achieves 100% accuracy in recognizing run, sit down, standup, and walk activities. However, there needs to be more clarification between lie down and sit down activities. The misclassification could be explained by the overlapping patterns between sit down activities, characterized by sudden sitting and prolonged immobility, and lie down activities.

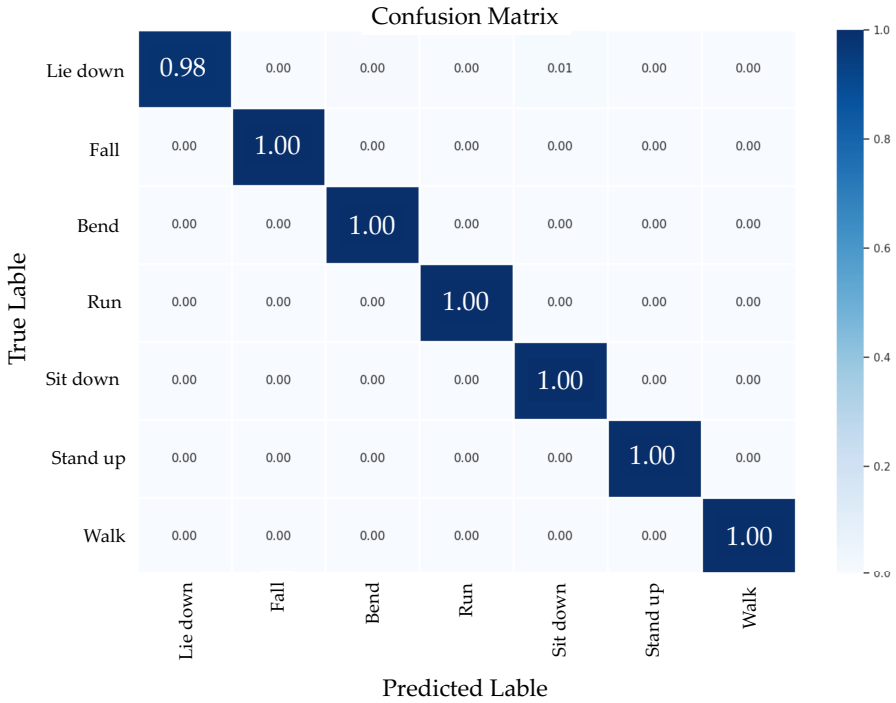

Figure 8. Confusion matrix of the proposed model on CSI-HAR dataset.

The accuracy and loss metrics of the CNN-GRU-AttNet model are illustrated in Figure 9. The graph in Figure 9a depicts the accuracy values for both the training and validation data. Notably, the model achieves convergence within a relatively short timeframe, specifically within 100 epochs. Additionally, Figure 9b demonstrates that the training loss exhibits higher values compared to the validation loss, which is a reasonable observation. This elevated training loss can be attributed to the multi-phase learning process aimed at understanding the distinct characteristics of CSI signals associated with various human actions.

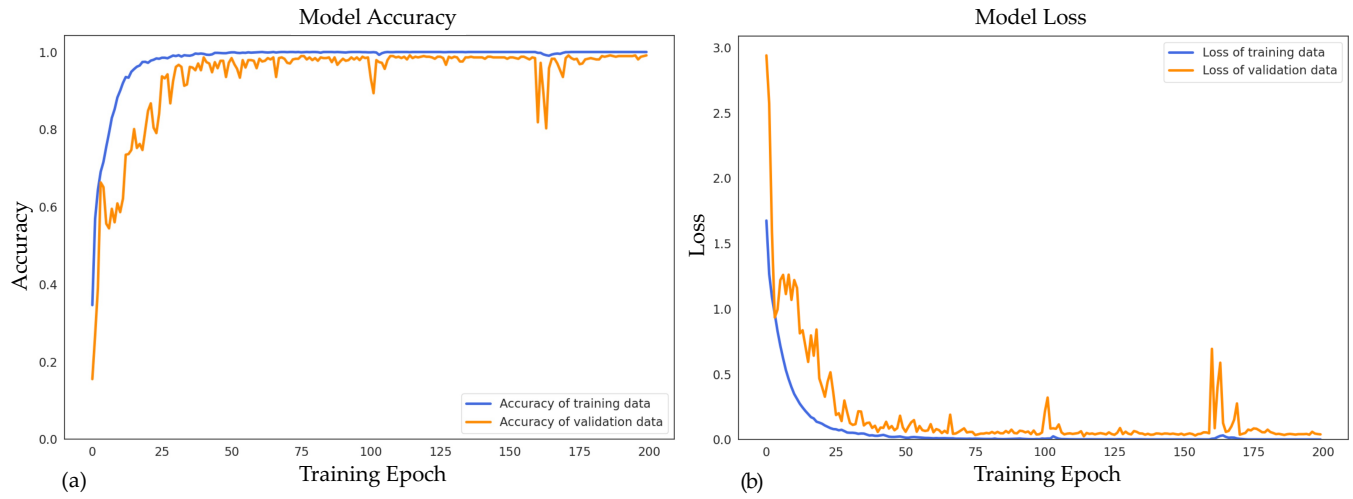

Figure 9. The accuracy and loss metrics of the CNN-GRU-AttNet model on the CSI-HAR dataset: (**a**) train and validation accuracy curves; (**b**) train and validation loss curves.

4.3. Experimental Findings on StanWiFi Dataset

Table 5 presents the experimental results of the CNN-GRU-AttNet model applied to HAR on the StanWiFi dataset. The table clearly demonstrates that the proposed model achieves impressive performance with an average accuracy of 98.66%, precision of 98.43%, recall of 97.88%, and F1-score of 98.14%. These results show that the CNN-GRU-AttNet model performs well in recognizing human activities. Furthermore, compared to five other baseline DL models, the CNN-GRU-AttNet model achieves the highest recognition accuracy.

Table 5. Performance results of both the proposed CNN-GRU-AttNet model and the five baseline models on the StanWiFi dataset.

Model	**Recognition Performance (Mean ± Std)**			
	Accuracy	**Precision**	**Recall**	**F1-Score**
CNN	89.08% (±4.61%)	87.52% (±5.47%)	89.49% (±3.48%)	87.55% (±4.74%)
LSTM	93.95% (±2.15%)	91.32% (±2.66%)	94.80% (±1.64%)	92.75% (±2.25%)
BiLSTM	94.73% (±1.73%)	92.25% (±2.15%)	94.74% (±1.18%)	93.18% (±1.77%)
GRU	94.84% (±2.52%)	92.68% (±3.44%)	94.84% (±2.32%)	93.35% (±3.12%)
BiGRU	95.73% (±2.64%)	94.70% (±2.39%)	95.02% (±3.78%)	94.62% (±3.38%)
CNN-GRU-AttNet	98.66% (±0.26%)	98.43% (±0.29%)	97.88% (±0.59%)	98.14% (±0.42%)

The confusion matrix of the CNN-GRU-AttNet model on the StanWiFi dataset is illustrated in Figure 10. The results demonstrate that the model has high recognition accuracy for lie down, walk, and stand up activities, achieving over 96% accuracy. In contrast, fall and run activities show slightly lower performance, around 94%. Additionally, sit down activities have a recognition accuracy of 92% or higher. However, there needs to be more clarification between lie down and sit down activities, and this may be due to the similarity in signal patterns between the two activities. Figure 11 illustrates the accuracy and loss metrics of the CNN-GRU-AttNet model applied to the StanWiFi dataset.

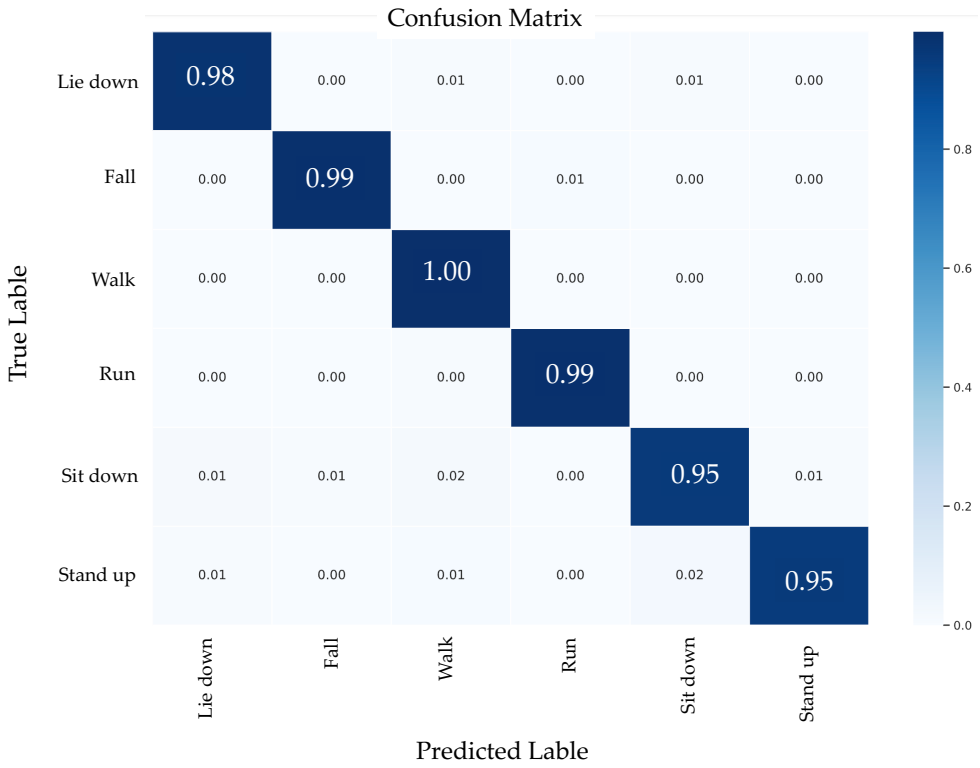

Figure 10. Confusion matrix of the proposed model on the StanWiFi dataset.

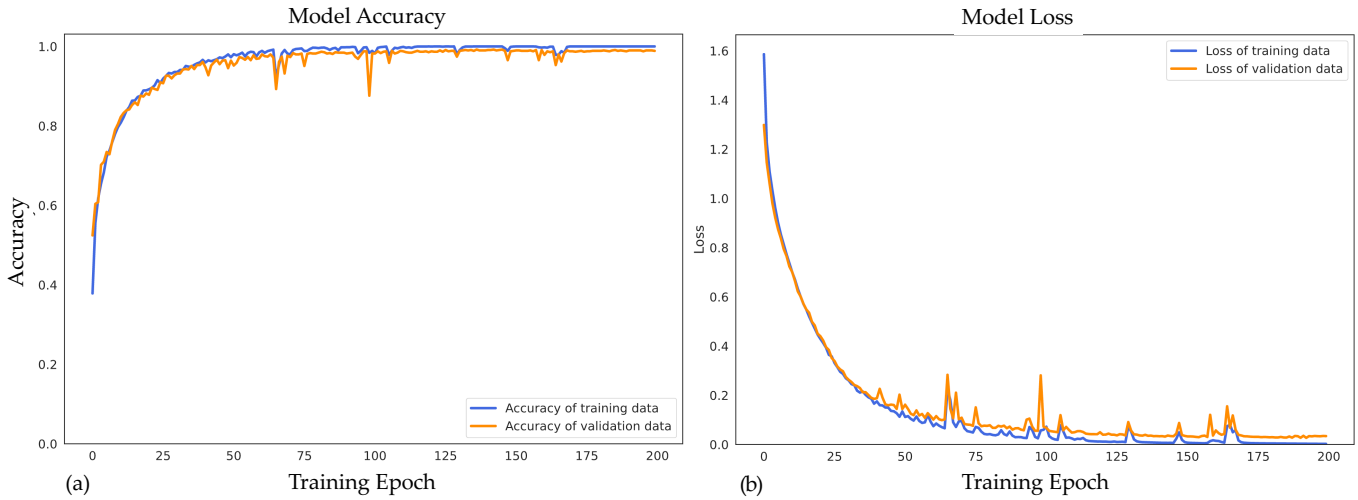

Figure 11. The accuracy and loss metrics of the CNN-GRU-AttNet model on the StanWiFi dataset: (**a**) train and validation accuracy curves; (**b**) train and validation loss curves.

5. Discussion

This section discusses the experimental outcomes achieved by utilizing the proposed CNN-GRU-AttNet model on two distinct datasets.

5.1. Performance Comparison

Assessing the overall effectiveness of a model presents a significant challenge, as it requires comparing different models using the same dataset. Therefore, we evaluate the efficacy of the proposed model through a comparative analysis with other models using the CSI-HAR and StanWiFi datasets. The comparative results are provided in Table 6. Our study demonstrates that the proposed CNN-GRU-AttNet model outperforms other models

on the CSI-HAR dataset in terms of recognition capability. The CNN-GRU-AttNet model achieved an average accuracy of 99.62%, precision of 99.61%, recall of 99.61%, and F1-score of 99.61%. Furthermore, our proposed CNN-GRU-AttNet model exhibits a remarkable accuracy improvement of 4.62% compared to the leading-edge model currently available on the CSI-HAR dataset.

The proposed CNN-GRU-AttNet model has demonstrated a superior level of accuracy, achieving a score of 98.66% on the StanWiFi dataset, which represents a 0.66% improvement over the best state-of-the-art performance [33]. Upon analyzing the performance of various existing models presented in Table 6, it is evident that our proposed CNN-GRU-AttNet outperforms all others in terms of recognition outcomes on both datasets. This observed enhancement can be attributed to the recently suggested framework, which adeptly captures and utilizes spatial and temporal characteristics extracted from unprocessed CSI data for the purpose of HAR.

Table 6. Comparison between the proposed model and other existing works.

Dataset	Classifier	Accuracy	Precision	Recall	F1-Score
CSI-HAR	2D-CNN [27]	95.00%	-	-	-
	MIMI-AE [28]	94.49%	-	-	-
	CNN-GRU-AttNet	99.62%	99.61%	99.61%	99.61%
StanWiFi	LSTM [19]	90.50%	-	-	-
	ABLSTM [20]	97.30%	-	-	-
	CSITime [33]	98.00%	-	-	-
	CNN-GRU-AttNet	98.66%	98.43%	97.88%	98.14%

5.2. Impact of the Attention Mechanism

The ability to obtain an interpretable representation is crucial for many ML applications. While DL techniques excel at extracting features from raw data, understanding the relative importance of the input data can be challenging. This issue has been addressed in prior research through the introduction of attention mechanisms. In our study, we enhanced the classification algorithm by incorporating an attention mechanism originally designed for neural network machine translation tasks, as presented by Luong et al. [37]. This approach allowed us to develop an interpretable representation that highlighted the significance of individual input data segments within the model. The findings of our study demonstrate that the inclusion of the attention mechanism led to improved recognition effectiveness across all scenarios, as evidenced by the data presented in Figure 12. The CNN-GRU-AttNet model exhibited notable performance improvements on both benchmark datasets.

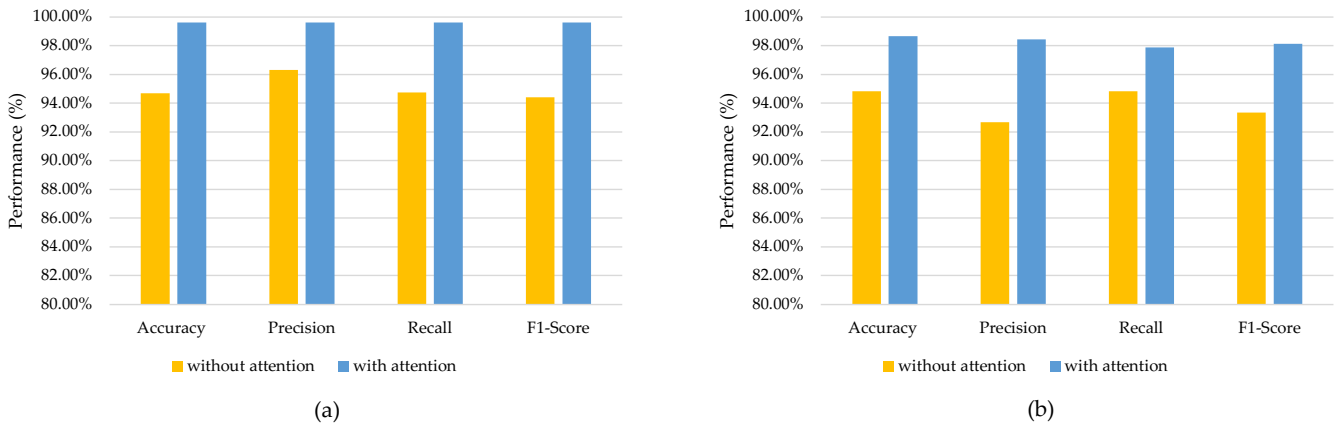

Figure 12. Improved performance of the proposed network with/without the attention mechanism: (**a**) CSI-HAR dataset; (**b**) StanWiFi dataset.

5.3. Impact of the PCA Denoising Method

In the proposed methodology, we employed the PCA denoising method to effectively eliminate noisy signals from the CSI data. Through experimentation, we observed that PCA leverages the correlated variations present in the CSI time series of different subcarriers, thereby effectively removing noise from the signals. This process specifically targets the elimination of uncorrelated noisy components that cannot be adequately filtered out through traditional low-pass filtering.

General-purpose denoising methods, such as low-pass filters or median filters, unfortunately, do not perform well in handling impulse and bursty noises for two reasons. Firstly, these methods typically require much higher sampling rates than the frequency of the WiFi signal, making them less suitable for this scenario. Secondly, the noise density in CSI values is too high for traditional filters to efficiently handle [34].

To thoroughly investigate the impact of the PCA denoising method, we conducted additional experiments. The comparative results presented in Table 7 clearly demonstrate that denoising the CSI data using PCA leads to notable improvements in the recognition performances of our proposed CNN-GRU-AttNet. Specifically, we achieved an accuracy increase of up to 1.43% for the CSI-HAR dataset and 1.00% for the StanWiFi dataset. As a result, it becomes evident that the PCA-based noise reduction plays a significant role in achieving the high recognition accuracies observed in our proposed methodology.

Table 7. Comparison between the proposed model using CSI data before and after denoising by the PCA denoising method.

Dataset	Classifier	Accuracy	Precision	Recall	F1-Score
CSI-HAR	CNN-GRU-AttNet using CSI data without the PCA denosing method	98.19%	98.27%	98.17%	98.17%
	CNN-GRU-AttNet using denoised CSI data with the PCA denosing method	99.62%	99.61%	99.61%	99.61%
StanWiFi	CNN-GRU-AttNet using CSI data without the PCA denosing method	97.66%	97.13%	97.00%	97.04%
	CNN-GRU-AttNet using denoised CSI data with the PCA denosing method	98.66%	98.43%	97.88%	98.14%

5.4. Performance Analysis for Different Subjects

To analyze performances across different subjects, we conducted an additional experiment using the CSI data from the CSI-HAR dataset, which includes detailed information about the subjects involved in data collection. Specifically, the CSI-HAR dataset contains records of each activity performed 20 times by 3 voluntary subjects of varying ages, ranging from 25 to 70 years old. The subjects represent a diverse group, consisting of an adult, a middle-aged person, and an elderly person.

Table 8 presents an analysis of the performance of the proposed CNN-GRU-AttNet model on individual subjects. Notably, the F1-scores of subject 2 (a middle-aged person) show consistently high values, exceeding 95% for all activities. On the other hand, when using the CSI data of subject 1 (an adult), the F1-scores for some activities (lie down, bend, sit down, stand up, and walk) are found to be lower than 95%. These findings strongly suggest that there are notable differences in the CSI data captured from different subjects.

5.5. Limitations of the Proposed Method

Due to the absence of a conventional dataset that incorporates the CSI data obtained from settings with substantial interference, we could not evaluate the suggested model's efficacy in such an environment. However, we aim to investigate this aspect in our future research.

Table 8. Recognition performances of the proposed CNN-GRU-AttNet based on CSI data from different subjects.

Subject	Activity	Recognition Performance		
		Accuracy	Recall	F1-Score
Subject 1 (an adult)	Lie down	82.4%	100.0%	90.3%
	Fall	100.0%	96.6%	98.2%
	Bend	86.2%	100.0%	92.6%
	Run	92.1%	100.0%	95.9%
	Sit down	100.0%	76.9%	87.0%
	Stand up	100.0%	87.0%	93.0%
	Walk	100.0%	89.3%	94.3%
Subject 2 (a middle-aged person)	Lie down	100.0%	96.3%	98.1%
	Fall	100.0%	100.0%	100.0%
	Bend	100.0%	100.0%	100.0%
	Run	100.0%	100.0%	100.0%
	Sit down	96.6%	96.6%	96.6%
	Stand up	100.0%	100.0%	100.0%
	Walk	96.7%	100.0%	98.3%
Subject 3 (an elderly person)	Lie down	100.0%	100.0%	100.0%
	Fall	100.0%	93.3%	96.6%
	Bend	88.9%	100.0%	94.1%
	Run	100.0%	95.5%	97.7%
	Sit down	100.0%	92.9%	96.3%
	Stand up	92.6%	100.0%	96.2%
	Walk	100.0%	100.0%	100.0%

6. Conclusions for Future Research

This study introduces a DL model called CNN-GRU-AttNet, designed to automatically recognize human behavior from WiFi CSI signals. Human activity can be represented as time-series data with temporal and spatial characteristics. The CNN-GRU-AttNet model addresses this challenge by extracting spatial and significant features simultaneously using convolutional blocks and attention modules, respectively. Additionally, the GRU block is employed to capture latent temporal patterns within the CSI signals. By combining these three components, the model effectively represents the CSI signal's characteristics and focuses its attention on activity-related information. Consequently, the CNN-GRU-AttNet model improves the accuracy of activity recognition. Evaluations were conducted on two distinct datasets, CSI-HAR and StanWiFi, resulting in recognition accuracies of 99.62% and 98.66%, respectively. A comparative analysis with existing approaches demonstrated the superiority of the proposed model, achieving improvements of 4.62% and 0.66% in accuracy, respectively.

There are potential plans to collect and analyze empirical datasets obtained from environments with significant interference. The task of identifying multi-user activity in real-world situations poses a more realistic and complex challenge compared to identifying single-user activity. As a result, this study's research will be expanded to include HAR for multiple users. Publicly available datasets often contain common activities that do not accurately represent real-world situations, as individuals engage in a variety of activities on a daily basis. Therefore, the acquisition of a WiFi dataset that includes a broader range of indoor human activities will be deferred for future research.

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

Funding: This project was funded by the Thailand Science Research and Innovation Fund; University of Phayao (Grant No. FF66-UoE001); National Science, Research and Innovation Fund (NSRF); and King Mongkut's University of Technology North Bangkok with Contract no. KMUTNB-FF-66-07.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data were presented in the main text.

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

References

1. Hnoohom, N.; Mekruksavanich, S.; Jitpattanakul, A. An Efficient ResNetSE Architecture for Smoking Activity Recognition from Smartwatch. *Intell. Autom. Soft Comput.* **2023**, *35*, 1245–1259. [CrossRef]
2. Thanarajan, T.; Alotaibi, Y.; Rajendran, S.; Nagappan, K. Improved wolf swarm optimization with deep-learning-based movement analysis and self-regulated human activity recognition. *AIMS Math.* **2023**, *8*, 12520–12539. [CrossRef]
3. Mekruksavanich, S.; Jitpattanakul, A. RNN-based deep learning for physical activity recognition using smartwatch sensors: A case study of simple and complex activity recognition. *Math. Biosci. Eng.* **2022**, *19*, 5671–5698. [CrossRef]
4. Jalal, A.; Kim, Y.H.; Kim, Y.J.; Kamal, S.; Kim, D. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognit.* **2017**, *61*, 295–308. .: 10.1016/j.patcog.2016.08.003. [CrossRef]
5. Sharma, V.; Gupta, M.; Pandey, A.K.; Mishra, D.; Kumar, A. A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets. *Appl. Artif. Intell.* **2022**, *36*, 2093705. [CrossRef]
6. Vrskova, R.; Kamencay, P.; Hudec, R.; Sykora, P. A New Deep-Learning Method for Human Activity Recognition. *Sensors* **2023**, *23*, 2816. [CrossRef]
7. 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]
8. Reyes-Ortiz, J.L.; Oneto, L.; Samà, A.; Parra, X.; Anguita, D. Transition-Aware Human Activity Recognition Using Smartphones. *Neurocomputing* **2016**, *171*, 754–767. .: 10.1016/j.neucom.2015.07.085. [CrossRef]
9. Mekruksavanich, S.; Hnoohom, N.; Jitpattanakul, A. A Hybrid Deep Residual Network for Efficient Transitional Activity Recognition Based on Wearable Sensors. *Appl. Sci.* **2022**, *12*, 4988. [CrossRef]
10. Yan, H.; Zhang, Y.; Wang, Y.; Xu, K. WiAct: A Passive WiFi-Based Human Activity Recognition System. *IEEE Sensors J.* **2020**, *20*, 296–305. [CrossRef]
11. Liu, F.; Cui, Y.; Masouros, C.; Xu, J.; Han, T.X.; Eldar, Y.C.; Buzzi, S. Integrated Sensing and Communications: Toward Dual-Functional Wireless Networks for 6G and Beyond. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 1728–1767. [CrossRef]
12. Wang, W.; Liu, A.X.; Shahzad, M.; Ling, K.; Lu, S. Device-Free Human Activity Recognition Using Commercial WiFi Devices. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1118–1131. [CrossRef]
13. Sigg, S.; Scholz, M.; Shi, S.; Ji, Y.; Beigl, M. RF-Sensing of Activities from Non-Cooperative Subjects in Device-Free Recognition Systems Using Ambient and Local Signals. *IEEE Trans. Mob. Comput.* **2014**, *13*, 907–920. [CrossRef]
14. Muaaz, M.; Chelli, A.; Pätzold, M. WiHAR: From Wi-Fi Channel State Information to Unobtrusive Human Activity Recognition. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; pp. 1–7. [CrossRef]
15. Cheng, X.; Huang, B. CSI-Based Human Continuous Activity Recognition Using GMM–HMM. *IEEE Sensors J.* **2022**, *22*, 18709–18717. [CrossRef]
16. Dang, X.; Cao, Y.; Hao, Z.; Liu, Y. WiGId: Indoor Group Identification with CSI-Based Random Forest. *Sensors* **2020**, *20*, 4607. [CrossRef] [PubMed]
17. Alsaify, B.A.; Almazari, M.M.; Alazrai, R.; Alouneh, S.; Daoud, M.I. A CSI-Based Multi-Environment Human Activity Recognition Framework. *Appl. Sci.* **2022**, *12*, 930. [CrossRef]
18. Moghaddam, M.G.; Shirehjini, A.A.N.; Shirmohammadi, S. A WiFi-based System for Recognizing Fine-grained Multiple-Subject Human Activities. In Proceedings of the 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Ottawa, ON, Canada, 16–19 May 2022; pp. 1–6. [CrossRef]
19. Yousefi, S.; Narui, H.; Dayal, S.; Ermon, S.; Valaee, S. A Survey on Behavior Recognition Using WiFi Channel State Information. *IEEE Commun. Mag.* **2017**, *55*, 98–104. [CrossRef]
20. Chen, Z.; Zhang, L.; Jiang, C.; Cao, Z.; Cui, W. WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM. *IEEE Trans. Mob. Comput.* **2019**, *18*, 2714–2724. [CrossRef]
21. Abdelnasser, H.; Youssef, M.; Harras, K.A. WiGest: A ubiquitous WiFi-based gesture recognition system. In Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM), Kowloon, Hong Kong, 26 April–1 May 2015; pp. 1472–1480. [CrossRef]
22. Gu, Y.; Quan, L.; Ren, F. WiFi-assisted human activity recognition. In Proceedings of the 2014 IEEE Asia Pacific Conference on Wireless and Mobile, Bali, Indonesia, 28–30 August 2014; pp. 60–65. [CrossRef]

23. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI: Indoor Localization via Channel Response. *ACM Comput. Surv.* **2013**, *46*, 1–32. [CrossRef]
24. Zhang, D.; Wang, H.; Wu, D. Toward Centimeter-Scale Human Activity Sensing with Wi-Fi Signals. *Computer* **2017**, *50*, 48–57. [CrossRef]
25. Wang, Y.; Liu, J.; Chen, Y.; Gruteser, M.; Yang, J.; Liu, H. E-Eyes: Device-Free Location-Oriented Activity Identification Using Fine-Grained WiFi Signatures. In Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom '14), New York, NY, USA, 2014; pp. 617–628. [CrossRef]
26. Wang, F.; Panev, S.; Dai, Z.; Han, J.; Huang, D. Can WiFi Estimate Person Pose? *arXiv* **2019**, arXiv:1904.00277.
27. Moshiri, F.; Parisa.; Shahbazian.; Reza.; Nabati.; Mohammad.; Ghorashi, S.A. A CSI-Based Human Activity Recognition Using Deep Learning. *Sensors* **2021**, *21*, 7225. [CrossRef]
28. Chahoushi, M.; Nabati, M.; Asvadi, R.; Ghorashi, S.A. CSI-Based Human Activity Recognition Using Multi-Input Multi-Output Autoencoder and Fine-Tuning. *Sensors* **2023**, *23*, 3591. [CrossRef] [PubMed]
29. Elbayad, M.; Besacier, L.; Verbeek, J. Pervasive Attention: 2D Convolutional Neural Networks for Sequence-to-Sequence Prediction. *arXiv* **2018**, arXiv:1808.03867.
30. Schuster, M.; Paliwal, K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
31. Zhang, J.; Wu, F.; Wei, B.; Zhang, Q.; Huang, H.; Shah, S.W.; Cheng, J. Data Augmentation and Dense-LSTM for Human Activity Recognition Using WiFi Signal. *IEEE Internet Things J.* **2021**, *8*, 4628–4641. [CrossRef]
32. Shang, S.; Luo, Q.; Zhao, J.; Xue, R.; Sun, W.; Bao, N. LSTM-CNN network for human activity recognition using WiFi CSI data. *J. Phys. Conf. Ser.* **2021**, *1883*, 012139. [CrossRef]
33. Yadav, S.K.; Sai, S.; Gundewar, A.; Rathore, H.; Tiwari, K.; Pandey, H.M.; Mathur, M. CSITime: Privacy-preserving human activity recognition using WiFi channel state information. *Neural Networks* **2022**, *146*, 11–21. [CrossRef] [PubMed]
34. Wang, W.; Liu, A.X.; Shahzad, M.; Ling, K.; Lu, S. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15), New York, NY, USA, 2015; pp. 65–76. [CrossRef]
35. andJimmy Ba, D.P.K. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–8 May 2015.
36. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 111–147. [CrossRef]
37. Luong, T.; Pham, H.; Manning, C.D. Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 1412–1421. [CrossRef]