This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

A Transformer-Based Methodology for Person-Independent Human Activity Recognition Using Wi-Fi CSI

Allan Costa Nascimento dos Santos[†,‡], Pamella Soares[§,‡], Iandra Galdino [†], Julio C. H. Soto[†], Cledson de Sousa[†], Taiane C. Ramos[†], Celio V. N. de Albuquerque[†], Raphael Guerra[†], Natalia C. Fernandes[†], Débora C. Muchaluat-Saade[†], Gheorghita Ghinea[‡]
[†]*MidiaCom Lab/Universidade Federal Fluminense – Niterói, RJ – Brazil*
[‡]*Department of Computer Science/Brunel University London – London, United Kingdom*
[§]*Graduate Program in Computer Science/ State University of Ceara – Ceara, Brazil*
{allans, igar, jsoto, taiane, celio, rguerra, natalia, debora}@midiacom.uff.br, george.ghinea@brunel.ac.uk
pamella.soares@aluno.uece.br

***Abstract*—The use of Channel State Information (CSI) for human activity recognition holds great promise in healthcare applications, particularly for remote patient monitoring. By interpreting variations in Wi-Fi signals, CSI can be leveraged to detect physical activities, falls, and daily movements. This capability enables the monitoring of patients without relying on wearable sensors or intrusive cameras, offering a fully non-invasive solution. Motivated by this potential, this paper proposes a wireless sensing model called `MDA-CSI` for recognizing human activities in indoor environments. `MDA-CSI` employs a Transformer-based architecture designed to process time-series information and effectively capture temporal dependencies. The proposed model is generalizable, allowing it to identify activities performed by individuals who were not included in the training phase.**

***Index Terms*—Transformer model, channel state information, Wi-Fi, wireless sensing system, proactive security, human activity recognition.**

I. Introduction

Human activity recognition for older adults, patients, people with disabilities, individuals with chronic diseases, or those in assisted treatment is increasingly important due to its impact on healthcare, quality of life, and well-being. Monitoring the physical activity of these populations helps identify sedentary behaviors and promotes the adoption of more active habits [1]–[3]. This can help prevent chronic diseases such as type-2 diabetes, heart disease, obesity, and osteoporosis, in addition to improving mental health and overall physical function [4], [5]. For individuals with disabilities or older adults recovering from injuries or surgeries, continuous monitoring of physical activity can provide valuable insights into rehabilitation progress. This enables healthcare professionals to tailor treatment plans and interventions based on accurate data regarding individual movement and mobility.

Regular physical activity among elderly individuals helps maintain independence for longer periods, reducing the need for daily assistance. Analyzing activity patterns can also help predict functional decline, allowing for early interventions. For people with disabilities, understanding movement and activity patterns supports the development of personalized exercise programs that enhance mobility, alleviate pain, and improve overall quality of life.

One of the major concerns for individuals with disabilities is safety, particularly regarding falls and medical emergencies [1], [6]. Sensors and monitoring systems can detect falls or extended periods of inactivity, automatically triggering alerts to caregivers or emergency services [1]. However, collecting physical activity data in uncontrolled environments often results in noisy or incomplete datasets, which makes accurate pattern interpretation challenging. Reliable analysis therefore requires robust data filtering techniques and a comprehensive understanding of movement contexts.

The cost of high-quality wearable devices and monitoring infrastructure can be a barrier to widespread implementation, especially among lower-income populations. Therefore, this paper proposes a Transformer-based methodology for Monitoring Daily Activity using Channel State Information (CSI), called `MDA-CSI`[1]. The `MDA-CSI` is a wireless sensing model aimed at overcoming the challenge of technology adoption among the elderly, as it does not require the use of wearable devices or any action from the patient. It analyzes the Wi-Fi signal transmitted by an access point, which is commonly present in almost all indoor environments. As an individual carries out daily activities, their movements result in changes in the signal received by `MDA-CSI`. By capturing and interpreting Wi-Fi signals in indoor environments, `MDA-CSI` can detect a patient's physical activity or daily movements and store this information, allowing caregivers and healthcare professionals to monitor the patient's condition without the need for wearable sensors or invasive cameras [5].

`MDA-CSI` uses a Transformer model [7], [8] designed to

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, CAPES Print, CNPq, FAPERJ, UFF and FINEP.

[1]https://github.com/mestrelan/MDA-CSI

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

process time series data. It incorporates positional encoding, multi-head attention, and a fully connected layer, enabling sophisticated learning of temporal relationships and generating accurate predictions based on patterns identified in the input sequence.

Unlike other machine learning models in the literature, `MDA-CSI` features a generalized approach capable of identifying the activities of individuals who did not participate in the model training phase. The system detects the presence of a person in a room and determines whether the person is moving. If the person is moving, the system further classifies the activity as walking or running. Otherwise, the system classifies the activity as lying down or sitting. `MDA-CSI` was trained, validated, and tested using CSI data from 59 volunteers of various genders, ages, and physical characteristics, collected in an indoor controlled environment [9], achieving an accuracy of 96.67%. We emphasize that the CSI dataset used in this study was originally collected and made publicly available by [9], and no new data collection was performed in this work. CSI data from 44 participants were used to train the model; data from 10 participants were used for validation, and data from 5 participants were used for testing.

The rest of the paper is organized as follows. Sec. II describes related work, while Sec. III presents the proposed methodology. The description of the experiments is provided in Sec. V. Sec. V-B discusses the results obtained. Finally, conclusions and suggestions for future work are presented in Sec. VI.

II. Related Work

Wi-Fi devices are now ubiquitous across a wide range of environments, with their signal characteristics influenced by various environmental factors, including human presence and movement [9], [10]. These variations can be captured in Channel State Information (CSI) data, which provide physical-layer (PHY) details such as amplitude and phase [1], [11], [12]. CSI shows strong potential for applications in elderly care [4], [10], as it enables unobtrusive monitoring in home or hospital settings without the need for substantial investment in dedicated sensing hardware [9].

Xiao et al. [13] proposed a semi-supervised generative adversarial network (GAN) for CSI-based activity recognition, called CsiGAN. The proposed method uses a complement generator, which can leverage limited unlabeled data to generate diverse fake samples for training a robust discriminator. For the discriminator, they modify the number of probability outputs, which helps establish a more accurate decision boundary. Additionally, they introduced manifold regularization to stabilize the learning process. The model was designed to address performance degradation in leave-one-subject-out validation for CSI-based activity recognition. However, the data used to train the model contained only a small number of candidates (3), with limited activities, resulting in low accuracy.

Caballero et al. [2] proposed an approach for human activity recognition (HAR) using commercial Wi-Fi devices. With their method, it is possible to infer the position of a monitored person in an indoor environment. To achieve this, they clean and process the amplitude of the CSI data collected from the Wi-Fi channel. For the scenario and dataset considered in this study, the results showed that the Random Forest (RF) algorithm performed the best in all tests, achieving an average accuracy of 93.03%. However, the RF model must be trained individually for each person to detect their position. Hence, the model must be trained on the individual's CSI data to function properly.

Wang et al. [14] proposed a multimodal channel state information-based activity recognition (MCBAR) system that leverages existing Wi-Fi infrastructures to monitor human activities using CSI measurements. MCBAR employs a multimodal generator to approximate the CSI data distribution across different environmental settings with limited measured CSI data. The generated CSI data from the multimodal generator provides diversity for knowledge transfer. A 2080ti GPU was used in their experiments. However, the model's training data were limited in both participant diversity and activity variety, restricting its applicability to new users.

Li et al. [15] proposed a human activity recognition scheme based on the collaboration between vision and Wi-Fi. They collected CSI data from Wi-Fi signals and human skeletal points from video. They developed a long-short-term Transformer network to combine the CSI data and skeleton points. The proposed method uses the Transformer neural network to form the long-short-term Transformer (LSTT) network, enabling it to derive the human skeleton points from the CSI data. Their method achieves 96% accuracy; however, it was tested using CSI data from only one person, limiting its ability to generalize to a broader population. Additionally, their method requires video data, which introduces significant limitations, including cost, data security concerns, and privacy issues.

Table I compares our proposal with other studies that use machine learning methods to monitor human activities through CSI. `MDA-CSI` stands out with 59 participants, a significantly higher number than that of most compared studies. This is important because a larger number of participants can increase data variability, leading to a more robust and generalizable model. The proposal is also notable for the quantity and variety of monitored activities. While the compared studies include between 5 and 6 activities, `MDA-CSI` covers 17 different postures and includes a condition with an empty room. This diversity enables the model to distinguish a broader range of behaviors and situations. The proposed system uses Transformer-based models, which are particularly effective in learning complex temporal patterns and may be better suited to capture the nuances of postural activities [16]. This gives it an advantage over traditional methods, such as GAN and Random Forest, used in the related work. Additionally, the proposed system is the only one that uses only real CSI data, independent of the person. This approach is advantageous for practical applications, as it eliminates the need for system calibration for specific individuals, thereby increasing its applicability in real-world scenarios. With an accuracy of 96.67%, the system

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

TABLE I: Comparative table of the proposal with related works.

Ref.	Participants	Activities	ML method	Uses only CSI real data independent of the person	Accuracy
[2]	125	Sitting, standing, lying, walking, running and sweeping	Random Forest	No	93.03%
[13]	3	Fall, walk, jump, pickup, sit down, and stand up	GAN	No	86.27%
[14]	10	Running, walking, falling down, boxing, circling arms, and cleaning floor	GAN	No	92.90%
[15]	3	Walking, waving hands, picking up, jumping, raising hands and squatting	Long-short-term Transformer	No	96%
MDA-CSI	**59**	**17 distinct postures plus one empty room collection**	**Transformer**	**Yes**	**96.67%**

is competitive with other methods. It is worth noting that it achieved this accuracy without relying on personalized data for each user. This combination of high accuracy and broad applicability demonstrates the efficiency and adaptability of the proposed system.

III. Methodology

This section presents the proposed methodology for human activity recognition. The MDA-CSI implements a training and evaluation flow of interactions to optimize accuracy, adjusting hyperparameters based on losses and recording metrics. The approach involves the development of four binary classification models, each addressing a specific aspect of human activity: presence, movement, posture (lying or sitting), and locomotion (running or walking).

A. CSI data collection

CSI data convey fundamental properties of communication channels, allowing us to describe how the signal changes as it propagates from the transmitter to the receiver.

In the IEEE 802.11ax specification [1], [2], [17], the physical layer of Wi-Fi networks utilizes the orthogonal frequency division multiplexing (OFDM) technique. OFDM is a multiplexing technique that divides the available bandwidth into several orthogonal subchannels [1], [4]. In this way, information can be transmitted independently on different subcarriers [2], [6], [18]. Furthermore, since subcarriers are orthogonal channels, each can provide unique CSI data; therefore, each subcarrier can be treated as an independent sensor capable of collecting CSI data [19].

In a Wi-Fi MIMO system under the IEEE 802.11n specification, with P transmission antennas and Q receiving antennas, the signal containing the estimated CSI data for each data stream can be expressed as follows:

$$h_{p,q} = |h|e^{j\theta}, \tag{1}$$

where $h_{p,q}$ represents the CSI between the p-th transmission antenna and the q-th receiving antenna, $|h|$ is the magnitude of the CSI signal, which is related to the signal attenuation during propagation, and $e^{j\theta}$ represents the phase of the CSI signal, which is related to the phase changes during propagation.

Since the channel is divided into multiple subcarriers in OFDM, the representation of the received signal will be a vector. Assuming that c is the number of subcarriers, the CSI between a pair of antennas (p, q) can be represented as a vector with c elements:

$$\boldsymbol{h}_{p,q} = [h_1, h_2, \ldots, h_c]^{\mathrm{T}}. \tag{2}$$

When a person is positioned between the transmitter and receiver, they act as an obstacle, affecting electromagnetic signal propagation. By analyzing these variations over time, human presence and body movements can be detected. The experiments were conducted in a 3m × 4m room containing three tables and a bed. The network consisted of a Wi-Fi router (transmitter), a laptop (Wi-Fi client and main packet receiver), and a Raspberry Pi 4B (CSI probe running NEXMON firmware). The participant sat at the center of the room, with the three devices placed 1m away: the router and laptop on opposite sides, and the Raspberry Pi equidistant from both. When in use, the bed was positioned 1m from the Raspberry Pi and aligned with the participant.

The CSI data used in this research [9] were collected from a Wi-Fi network operating at 5 GHz with an 80 MHz bandwidth. To ensure high data quality, the experimental setup included spectrum analysis to identify an unoccupied channel within the 5 GHz ISM band. The laptop continuously sent pings to the router to generate traffic, while a Raspberry Pi 4B with a single antenna and NEXMON firmware passively captured CSI at 33–34 samples per second over 60 s, yielding approximately $2,000$ samples per minute across 256 subcarriers. The dataset comprises 17 distinct postures, plus an empty-room condition, with one-minute recordings for each activity—covering a broader range of human behaviors than previous studies. This study was approved by the Ethics Committee of Fluminense Federal University (CAAE 54359221.4.0000.5243).

B. CSI data preprocessing

Fig. 1 shows the variability of subcarriers across an 80 MHz channel centered at 5 GHz. The results indicate a marked increase in variance among the first 60 subcarriers, with the remaining subcarriers demonstrating relatively stable characteristics. For our experiments, we restrict the analysis to frequencies below subcarrier 60, excluding any null or pilot subcarriers within this range. Thus, we utilized the first 60 subcarriers, resulting in a matrix of 2000 samples by 60 subcarriers for each data collection (activity type) from each participant. After that, we removed 12 null and pilot subcarriers, which do not carry meaningful data, leaving us with 48 subcarriers. Therefore, in OFDM, not all subcarriers

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

have the same signal strength. The first 40 subcarriers concentrate most of the signal energy. Therefore, there is greater amplitude/phase variation in them when a person is present (more sensitivity to movement). If another channel were used, the energy distribution pattern might change slightly, but the tendency for the initial subcarriers to carry more useful information would remain (this is a physical characteristic of Wi-Fi OFDM).

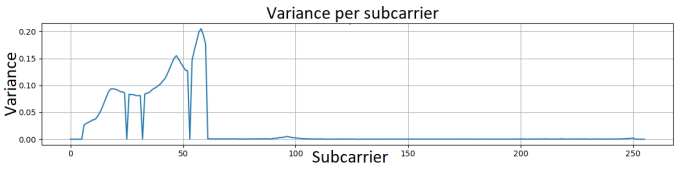


Fig. 1: Variance per subcarrier

Subsequently, we calculated the amplitudes of the signals collected within the subcarriers. This calculation was performed using the components of the complex numbers obtained from the collected signals. Since a complex number is defined by its real and imaginary parts, we determined the amplitude based on the magnitude derived from these components. Thus, we obtained amplitudes for all 2000 samples across the 48 resulting subcarriers for each data collection. The CSI data was then processed into a one-minute-long time series (the collection duration of one activity from the dataset), which was used as input to the Transformer model [8].

In our dataset, we used CSI data from 59 volunteers of various genders, ages, and physical characteristics. CSI data from 44 participants were used to train the model; data from 10 participants were used for validation, and data from 5 participants were used for testing. In the CSI processing, the null and pilot subcarriers were removed, leaving only the 48 active subcarriers, which form the OFDM subblock developed by the model. This means that, at each instant of time, the model obtains information from 48 distinct frequencies, allowing it to capture variations in the communication channel over time. An OFDM subblock is a frequency spectrum subdivision used in wireless communication to transmit data efficiently and robustly against interference and channel distortion. In the context of CSI, subcarriers represent individual frequencies within an OFDM block [11], [17]. Each subcarrier can be seen as a small transmission channel within the total frequency band, allowing multiple data to be sent simultaneously.

C. Transformer model

Fig. 2 shows the Transformer Encoder layers. Data Input to the Encoder: The data is entered in the format [batch_size, seq_len, num_channels], where batch_size is the number of samples processed simultaneously, seq_len is the number of instants in time and num_channels is the number of variables measured at each instant.

The encoder consists of several identical stacked layers, as determined by the parameter num_layers, where each layer follows a series of steps. First, the Multi-Head Self Attention Mechanism [7] enables the model to learn relationships between different moments in the sequence. Each attention head, defined by num_heads, processes a distinct part of the information.

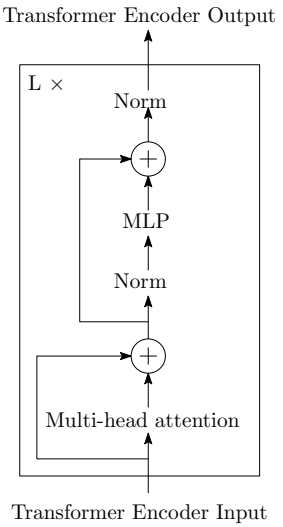


Fig. 2: Transformer Encoder Layers.

Summarizing the steps of the process, the input of the data with positional encoding, followed by Multi-head Self-Attention to find temporal dependencies, then Normalization and residual connections for stabilization, then the Feed-Forward Network to capture more complex patterns, more normalization and residual connection and repetition of the process num_layers times. This structure allows the Transformer to learn global and local patterns in time series, without the limitation of recurrence (as in RNNs) with highlighted temporal relationships [7], [8].

`MDA-CSI` implements a sequence of tests based on different hyperparameters. Each combination is evaluated on the validation set according to the metrics of accuracy, recall, precision and F1-score and the final model is evaluated on the test dataset. `MDA-CSI` explores the optimization of a Transformer model for binary classification of CSI time series, evaluating different combinations of hyperparameters. The goal is to identify the configurations that maximize performance in terms of accuracy, precision, recall and F1. The focus is to explore how the variables number of heads, number of layers, learning rate, number of epochs and batch size affect performance metrics. For this purpose, the `MDA-CSI` was developed to run the following hyperparameters: **Number of heads**: [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]; **Number of layers**: [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]; **Learning rate** (*lr*): [0.00001, 0.0001, 0.001, 0.01] **Number of epochs**: [20, 40, 60, 100]; **Batch size**: [4, 8, 12, 16, 20, 24]. Table II shows the hyperparameter combinations of the final model.

The Heads and Layers columns in the Table II specify the number of attention heads and Transformer layers for each task. In general, more complex tasks, such as sit/lying detection, use more heads (8) and intermediate layers (4), indicating the need for greater attention capacity to capture

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

TABLE II: The parameter combinations of the final models.

Activity	Heads	Layers	L. Rate	Epochs	Batch Size
Presence D.	3	6	0.0001	60	24
Movement D.	3	2	0.0001	100	4
Walk/Run D.	3	11	0.00001	100	24
Sit/Lying D.	8	4	0.0001	20	4

nuances in postures. On the other hand, simpler tasks, such as presence and motion detection, use a lighter configuration, with 3 heads and 6 and 2 layers. The learning rate is adjusted for each task.

D. Model training

Our main goal is to recognize human physical activities, for which we developed four independent models that take as input the previously described collected and processed CSI data. The models were developed to detect and classify a variety of human activities within a room, including presence detection, movement detection, gait classification (walking vs. running), and posture classification (sitting vs. lying down). Each of the generated models uses specific positions from the dataset. These models perform binary classification, yielding only two possible outputs.

For the generation and training of Model 1 (presence detection), data segmentation of the CSI data was necessary. This involved selecting data from empty rooms and rooms with human presence. Additional collection was used for empty rooms, while data from all 17 activities were used for rooms with human presence. The data were labeled as: no presence for empty rooms and presence for rooms with human presence.

E. Validation and testing

The validation set is used to evaluate the performance of the trained models by computing loss and accuracy metrics without altering the model's parameters. To ensure deterministic evaluation, specific training behaviors like Dropout and Batch Normalization are disabled, and gradient calculations are halted. The model iterates over validation batches, processing input data and ground truth labels, and accumulates loss and accuracy metrics. We selected the model that achieved the highest accuracy and lowest loss during the validation phase for use in the testing stage. The test evaluates the performance of the trained model on never-seen data, where we calculate various metrics, including loss, accuracy, precision, recall, F1-Score, and a confusion matrix.

IV. Implementation

Fig. 3 shows the overall processing flow of the MDA-CSI structure, organized to follow the execution and data processing flow. A detailed explanation of each stage of the process follows.

Each step includes the corresponding function and its purpose. MDA-CSI implements and trains a Deep Learning Transformer model for a binary classification task, where a sequence of 2000 samples for each carrier is classified into one of two categories. As a result, only one label is required

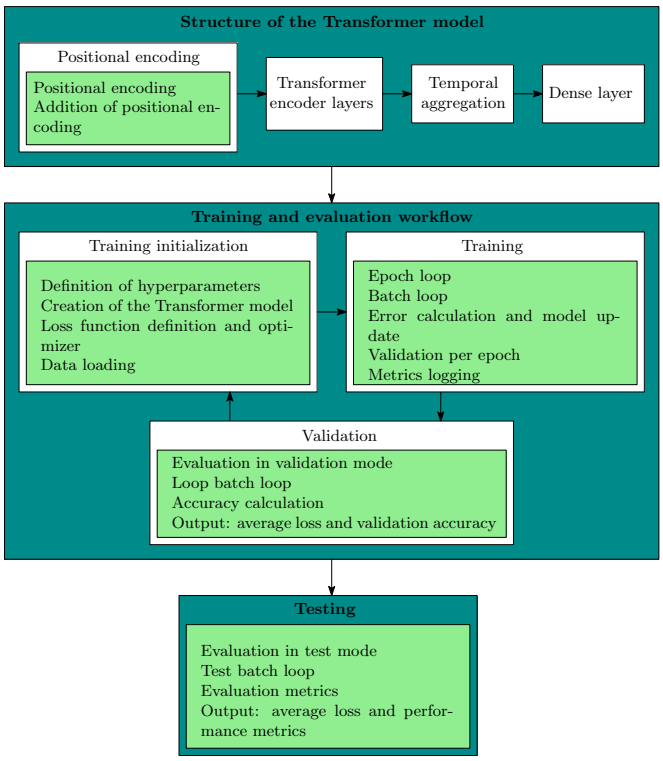


Fig. 3: Block diagram of the MDA-CSI.

for each entire time series. The initialization begins by defining the hyperparameters and devices, specifying parameters such as the number of heads, layers, learning rate, and whether training will occur on the CPU or GPU.

The Transformer model structure includes positional encoding, which is applied to add position information to the input data. The data is then passed through the Transformer encoder, which applies multi-head self-attention and feed-forward layers multiple times (as defined by the number of layers) to learn temporal dependencies. Temporal aggregation is performed by calculating the average of the representations along the temporal dimension, summarizing the sequence information. Finally, a fully connected layer maps the aggregated output to the final prediction, adjusting the model for binary classification.

V. Experiments

In the following sections, we present the setup used for the experiments conducted and the results obtained. Additionally, we compare the results with those found in the literature.

A. Setup

Regarding to the group of participants, a total of 59 volunteers, comprising both males and females, were recruited for the experiments. Out of the total, 44 participants (75%) were used for model training, 10 participants (15%) for validation, and 5 participants (10%) for testing. Fig. 4 presents the demographic characteristics of all participants, including age, height, and weight distributions as follows. As shown in the

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

histograms, the *mean age* of participants was 26.6 years, while the *mean height* was 172 cm and the *mean weight* was 73 kg.

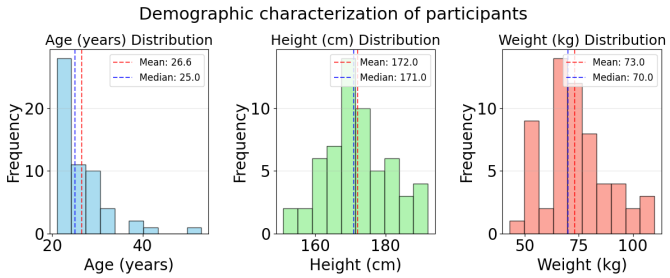


Fig. 4: Demographic characterization of participants.

Concerning the computational resources, we conducted the tests using three machines with the following configurations and using Anaconda to manage all tools, with Python v3.8.18, numpy 1.24.4, pandas 2.0.3 and pytorch 2.2.0: **Machine 1:** Windows 10 64 bits, CPU Intel(R) 12th Gen Core(TM) i5-12400F - 2.50 GHz, 12 GB RAM, SSD 250 GB, NVIDIA GeForce GTX 1660 SUPER graphics card. **Machine 2:** Ubuntu 22.04.4 LTS 64-bit, 4 TB of disk capacity, NVIDIA GeForce GTX 1660 SUPER graphics card, Intel Core i7-10700F CPU @ 2.90 GHz x 16, 16 GB RAM. **Machine 3:** Debian 12 64-bit, 1 TB of disk space, NVIDIA GeForce GTX 1660 graphics card, 11th Gen Intel Core i7-11700F x 16 processor, 16 GB RAM.

As detailed in Section III-A, specific configurations were employed for data collection, which took place in a dedicated room at the Fluminense Federal University's Computer Science Institute. Data was collected using a Raspberry Pi B4 equipped with a bcm43455c0 chipset. The experiments ensured a direct line of sight between the devices involved, positioned approximately one meter from the participants, as illustrated in the Fig. 5.

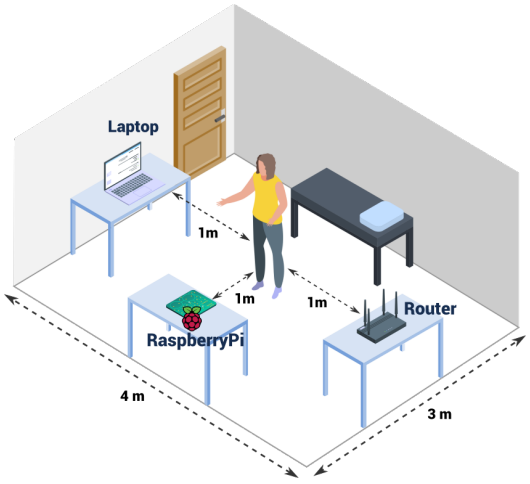


Fig. 5: Illustration of the experiment room at the Computer Science Institute of the Fluminense Federal University.

Furthermore, participants were not restricted in their use of clothing or electronic devices, and the experimental environment was designed to closely mimic a domestic setting represented by 17 activities, as illustrated in the Fig. 6.

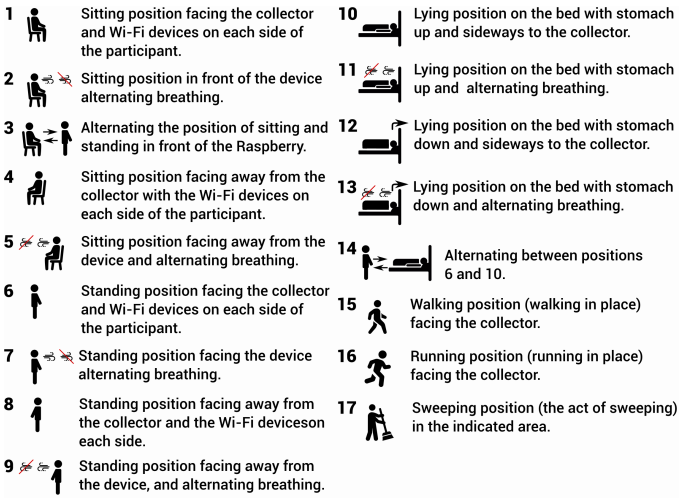


Fig. 6: All the 17 activities performed by volunteers.

B. Results

Fig. 7 presents the confusion matrices of the test results of each of the four models described previously, namely (a) presence detection, (b) movement detection, (c) walking or running activity recognition, and (d) sitting or lying posture recognition. Fig. 7 reflects a balanced test set, composed of 5 volunteers. Due to the balance of the data, the presence detection and walking/running tasks present confusion matrices with 5 test cases, representing one test per volunteer. This is due to each volunteer contributing with a specific sample for these activities. For motion and posture detection tasks (such as sitting/lying), the test set includes more samples, allowing the model to perform multiple tests per volunteer. This increases the validation and performance analysis capabilities since the model can capture a wider variety of instances for each volunteer, providing a more comprehensive assessment of performance in complex activities.

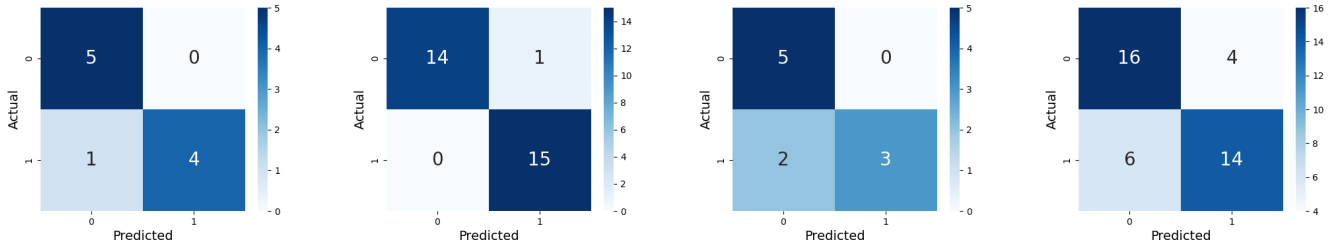
Table III presents the overall results obtained from identifying different activities of the test volunteers. The analysis starts with human presence detection in a room. The model achieved an accuracy of 90% and a perfect precision of 100%, demonstrating that all instances of human presence were correctly identified. The F1-Score further reinforces the robustness of the results.

TABLE III: General test results.

Activity	Acc.	Precision	Recall	F1-sc.	Test Loss
Presence D.	90%	100%	80%	0.89	0.56674
Movement D.	96.67%	93.75%	100%	0.97	0.15423
Walk/Run D.	80%	100%	60%	0.75	0.69235
Sit/Lying D.	75%	77.78%	70%	0.74	0.95841

The `MDA-CSI` movement detection model achieved an accuracy of just over 96%. The sensitivity, or recall, which measures the true positive rate, was perfect 100%, with a

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)



(a) Presence c. matrix (b) Movement c. matrix (c) Walking/Running c. matrix (d) Sitting/Lying c. matrix

Fig. 7: Confusion matrices for different activity recognition tasks.

corresponding F1-Score of 0.97. These results indicate that the model is highly effective in detecting human motion within an enclosed environment. When combined with presence identification, this technology has the potential for application in various non-critical everyday scenarios.

When detecting specific activities such as walking/running, and sitting/lying down, the model achieved accuracies of 80% and 75%, respectively. These results offer a promising foundation for future research and improvements to the model. It is important to note that detecting activities with high levels of movement, like running and walking, presents a significant challenge. The inherent nature of these movements introduces noise into the CSI data, making accurate detection more difficult. For Walking/Running activity detection, the F1-score of 0.75 performed more modestly. The recall of 60% suggests that it struggles to capture all cases, leaving room for future improvements. In Sitting/Lying down position detection, it achieved a precision of 77.78% and a recall of 70%. The F1-Score of 0.74 shows reasonable performance, although there is also room for improvement, especially in recall.

Test loss values vary significantly across activities, with motion detection exhibiting the lowest loss and sitting/lying detection showing the highest. This indicates that the model performs more efficiently on simpler activities or those with clearer patterns, while tasks involving greater variability or postural complexity, such as sitting/lying and walking/running, require further adjustments to reduce test loss and improve generalization.

Fig. 9 shows the train loss over the epochs. Loss is a metric used to determine how well the model is performing in training, with lower values indicating better performance. The training loss value appears to decrease overall, which is a good sign as it indicates that the model is learning and improving over time. The initial loss value of approximately 0.8288 decreases to approximately 0.6442. Fig. 8 shows the validation loss and accuracies over the epochs. The accuracies on the validation set represent the proportion of correct predictions made by the model. Higher rates indicate a more effective model. During the validation of the model for presence detection, Fig. 8 shows that the initial loss presented fluctuations, stabilizing and reaching lower values after epoch 40. This pattern was repeated in the accuracy, which also stabilized after epoch 40. Although the model demonstrated slower learning, it finally converged to satisfactory results. Fig. 8 shows that this initial difficulty may be related to the similarity between the CSI data of empty rooms. In contrast to presence detection, motion detection presents faster and more stable learning, reaching desirable loss and accuracy values before epoch 20. This superior performance can be attributed to the greater distinction between the CSI data of the participants present in the room.

Table IV presents the performance metrics of the GPUs employed in the experiments. In most cases, the GPU utilization was consistently high, typically around 98-100%. This indicates efficient utilization of the GPU resources. The power consumption varied across different activities and machines, ranging from 49.8W to 100.27W. The memory usage was generally higher for more complex tasks like movement detection, reaching up to 5853MB. Training time varied significantly depending on the complexity of the task and the number of epochs. For instance, training the model for presence detection took significantly less time than training it for movement detection. Testing time was low for all activities, indicating efficient inference. The models performed very well in total testing time (total time to infer the activity of all participants in the test set) even when using CPU. It is important to mention that the models can be trained using a GPU or CPU, with the training weights saved to be used for testing on a GPU or CPU, offering greater adaptation to different contexts.

By analyzing GPU metrics such as training and testing time, power consumption, utilization, temperature, and memory used, we are gaining valuable insights into the efficiency and practical viability of machine learning models. In the healthcare context, many systems need to operate consistently and reliably, often on remote monitoring devices or on local servers. The GPU power draw (W) metric provides insight into the energy cost, which is essential for optimizing monitoring systems and keeping operational costs low. Temperature (C) is crucial for the durability of hardware, especially in environments where it needs to operate continuously. The GPU Utilization (%) and Memory used (MB) metrics indicate how well the hardware is being utilized. High sustained utilization suggests the need for more powerful hardware to handle larger or more complex data. Total training time and Total testing time reflect the time it takes to get a solution up and running and its ability to adapt to new data. In physical activity

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

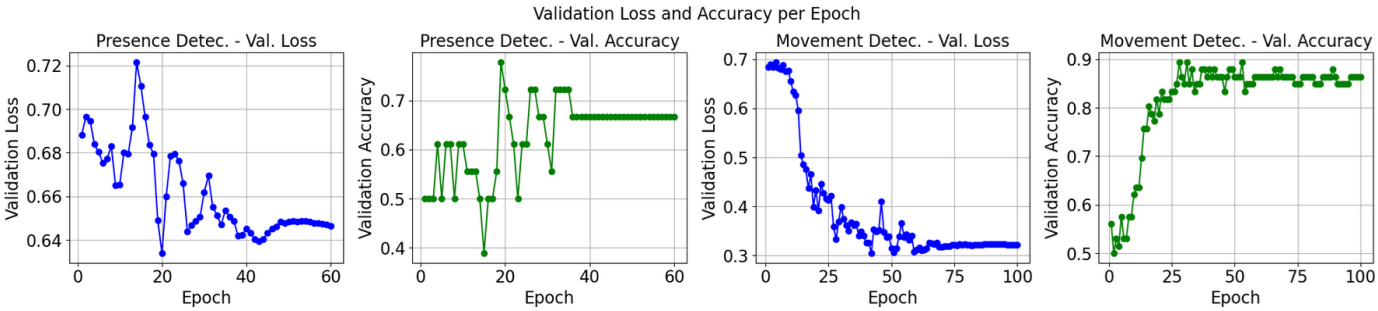


Fig. 8: Validation Loss and Accuracy per Epoch of the MDA-CSI.

TABLE IV: GPU utilization results.

Activity	Machine	Epochs	Total Train Time(s)	Total Test Time(s)	Power Draw(W)	Util.(%)	Temp.(°C)	Memo. Used(MB)
Pres. Detec.	2(CPU)	20	765.76	0.7475	12.93	0	34	173
Pres. Detec.	2	40	74.69	0,1217	74.50	99	54	3066
Pres. Detec.	1	20	200.07	2.6280	66.13	100	60	5604
Mov. Detec.	1	100	2875.12	2.8232	62.40	99	55	5592
Mov. Detec.	3	100	436.99	0.1471	100.27	93	81	2700
Mov. Detec.	2	100	451.43	0.1553	52.88	98	56	2688
Mov. Detec.	2(CPU)	20	2315.09	2.0087	12.51	0	34	167
Walk/Run. Detec.	2	100	125.06	0.0642	94.06	92	63	2683
Walk/Run. Detec.	2(CPU)	20	823.00	0.7363	112.04	0	34	142
Walk/Run. Detec.	1	100	1588.62	2.9641	49.8	98	57	5853
Sit./Lying Detec.	2	20	1110.41	1.6788	83.08	98	56	5606
Sit./Lying Detec.	2(CPU)	20	3127.92	2.5642	12.09	0	34	123
Sit./Lying Detec.	1	20	1433.80	3.9714	67.62	99	52	5807

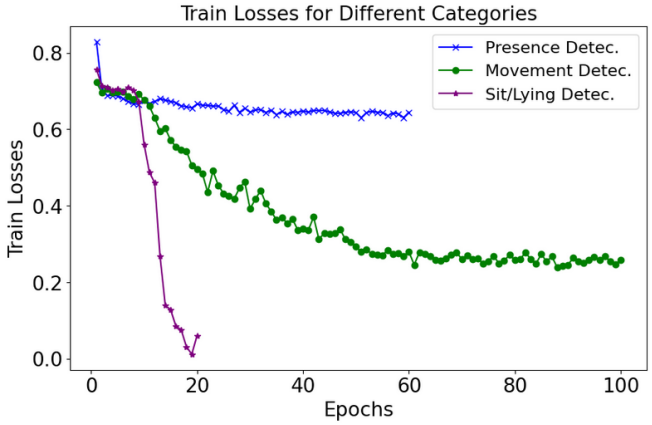


Fig. 9: Train Losses for Different Categories of the MDA-CSI.

monitoring, where older adults' behavioral patterns can change over time, rapid training and responsiveness ensure that model remains relevant and accurate.

C. Comparison with related work

CsiGAN [13] used a dataset (FallDeFi Data) with 3 volunteers, and MCBAR [14] with 10 volunteers. The dataset used by MDA-CSI contains CSI data from 59 volunteers, ensuring greater variability. MDA-CSI has greater movement identification accuracy (96.67%) than the proposals presented in [13] (86.27%), in [14] (92.90%), and in [2] (93.03%).Although the MDA-CSI performance is similar to that obtained in [2], the proposed approach has the advantage of being general enough to recognize the activities of people not included in the training dataset. Furthermore, unlike other related work, MDA-CSI identifies a person's activity in real time, taking approximately 0.56 s to process.

VI. Conclusions

This paper introduces MDA-CSI, a methodology designed to recognize human activities in indoor environments through Channel State Information (CSI) data analysis. MDA-CSI improves the accuracy of human activity recognition, even for individuals excluded from the training phase. Trained and evaluated on a dataset comprising 59 volunteers, MDA-CSI achieved an accuracy of 96.67% in movement detection, demonstrating strong generalization across diverse subjects. The motion detection model exhibited fast and stable convergence, reaching optimal loss and accuracy levels with relatively few training epochs.

Key strengths of MDA-CSI include the large number of participants involved in the experiments, the diversity of monitored activities, the adoption of Transformer-based models, and its user-independent design. Collectively, these features make MDA-CSI a robust and scalable framework for real-world human activity monitoring. The results are particularly encouraging for applications in elderly care, where non-invasive observation of daily activities is crucial. Moreover, MDA-CSI 's ability to leverage existing Wi-Fi infrastructure provides a cost-effective alternative to specialized sensing technologies. By employing a fully non-intrusive monitoring approach, MDA-CSI removes the need for physical contact or wearable devices, ensuring both user comfort and ease of deployment.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/WD67713.2025.11302651, 2025 13th Wireless Days Conference (WD)

REFERENCES

[1] J. C. Soto, I. Galdino, E. Caballero, V. Ferreira, D. Muchaluat-Saade, C. Albuquerque, A survey on vital signs monitoring based on wi-fi csi data, Computer Communications 195 (2022) 99–110.

[2] E. Caballero, I. Galdino, J. C. Soto, T. C. Ramos, R. Guerra, D. Muchaluat-Saade, C. Albuquerque, Human activity recognition using wi-fi csi, in: International Conference on Pervasive Computing Technologies for Healthcare, Springer, 2023, pp. 309–321.

[3] A. C. N. dos Santos, K. de Paula, M. T. L. Vidal, J. M. M. da Silva, C. de Sousa, L. A. F. Fernandes, T. B. de Castro, M. Bedo, T. C. Kohwalter, C. A. M. Bastos, F. L. Seixas, N. C. Fernandes, D. C. Muchaluat-Saade, G. Ghinea, A computer vision model to support individuals with disabilities within university campuses, in: 2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom), 2024, pp. 1–7. doi:10.1109/HealthCom60970.2024.10880838.

[4] J. C. H. Soto, I. Galdino, B. G. Gouveia, E. Caballero, V. Ferreira, D. Muchaluat-Saade, C. Albuquerque, Wi-fi csi-based human presence detection using dtw features and machine learning, in: 2022 IEEE Latin-American Conference on Communications (LATINCOM), 2022, p. 1.

[5] B. G. Gouveia, I. Galdino, E. Caballero, J. C. H. Soto, T. C. Ramos, R. Guerra, D. Muchaluat-Saade, C. V. N. Albuquerque, Parameter tuning for accurate heart rate measurement using wi-fi signals, in: 2024 International Conference on Computing, Networking and Communications (ICNC), 2024, pp. 407–411. doi:10.1109/ICNC59896.2024.10556168.

[6] Z. Chen, L. Zhang, C. Jiang, Z. Cao, W. Cui, Wifi csi based passive human activity recognition using attention based blstm, IEEE Transactions on Mobile Computing 18 (11) (2019) 2714–2724.

[7] D. Rothman, Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more, Packt Publishing Ltd, 2021.

[8] M. Joseph, Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning, Packt Publishing Ltd, 2022.

[9] I. Galdino, J. C. H. Soto, E. Caballero, V. Ferreira, T. C. Ramos, C. Albuquerque, D. C. Muchaluat-Saade, ehealth csi: A wi-fi csi dataset of human activities, IEEE Access 11 (2023) 71003–71012.

[10] Y. Wang, K. Wu, L. M. Ni, Wifall: Device-free fall detection by wireless networks, IEEE Transactions on Mobile Computing 16 (2) (2017) 581.

[11] C. Chen, G. Zhou, Y. Lin, Cross-domain wifi sensing with channel state information: A survey, ACM Computing Surveys 55 (11) (2023) 1–37.

[12] A. A. Milan, N. C. Fernandes, D. S. V. Medeiros, A monte carlo approach for antenna blocking probability estimation in mobile networks, in: 2022 25th Conference on Innovation in Clouds, Internet and Networks (ICIN), 2022, pp. 146–150. doi:10.1109/ICIN53892.2022.9758117.

[13] C. Xiao, D. Han, Y. Ma, Z. Qin, Csigan: Robust channel state information-based activity recognition with gans, IEEE Internet of Things Journal 6 (6) (2019) 10191–10204.

[14] D. Wang, J. Yang, W. Cui, L. Xie, S. Sun, Multimodal csi-based human activity recognition using gans, IEEE Internet of Things Journal 8 (24) (2021) 17345–17355.

[15] S. Li, Y. Ge, M. Shentu, S. Zhu, M. Imran, Q. Abbasi, J. Cooper, Human activity recognition based on collaboration of vision and wifi signals, in: 2021 International Conference on UK-China Emerging Technologies (UCET), 2021, pp. 204–208.

[16] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, R. P. Ramachandran, G. Rasool, Transformers in time-series analysis: A tutorial, Circuits, Systems, and Signal Processing 42 (12) (2023) 7433–7466.

[17] Y. Ma, G. Zhou, S. Wang, Wifi sensing with channel state information: A survey, ACM Computing Surveys (CSUR) 52 (3) (2019) 1–36.

[18] S. Weinstein, P. Ebert, Data transmission by Frequency-Division Multiplexing using the Discrete Fourier Transform, IEEE Transactions on Communication Technology 19 (5) (1971) 628–634.

[19] S. Lee, Y. D. Park, Y. J. Suh, S. Jeon, Design and implementation of monitoring system for breathing and heart rate pattern using WiFi signals, IEEE Annual Consumer Communications and Networking Conference (2018) 1–7.