



# CSI-Based Location Independent Human Activity Recognition Using Deep Learning

Fahd Abuhoureyah<sup>1</sup> · Yan Chiew Wong<sup>1</sup> · Ahmad Sadhiqin Bin Mohd Isira<sup>1</sup> · Mohammed Nasser Al-Andoli<sup>2</sup>

Received: 20 April 2023 / Accepted: 19 September 2023  
© The Author(s) 2023

## Abstract

Human Activity Recognition (HAR) is widely used in various applications, from smart homes and healthcare to the Internet of Things (IoT) and virtual reality gaming. However, existing HAR technologies suffer from limitations such as location dependency, sensitivity to noise and interference, and lack of flexibility in recognizing diverse activities and environments. In this paper, we present a novel approach to HAR that addresses these challenges and enables real-time classification and absolute location-independent sensing. The approach is based on an adaptive algorithm that leverages sequential learning activity features to simplify the recognition process and accommodate variations in human activities across different people and environments by extracting the features that match the signal with the surroundings. We employ the Raspberry Pi 4 and Channel State Information (CSI) data to extract activity recognition data, which provides reliable and high-quality signal information. We propose a signal segmentation method using the Long Short-Term Memory (LSTM) algorithm to accurately determine the start and endpoint of human activities. Our experiments show that our approach achieves a high accuracy of up to 97% in recognizing eight activities and mapping activities associated with environments that were not used in training. The approach represents a significant advancement in HAR technology and has the potential to revolutionize many domains, including healthcare, smart homes, and IoT.

**Keywords** Human activity recognition · Deep learning · Location-independent sensing · Nexmon CSI based

## Abbreviations

HAR	Human activity recognition
IoT	Internet of things
RPi	Raspberry Pi
CSI	Channel state information
LSTM	Long short-term memory
CNN	Convolutional neural network
BVP	Body velocity profile
MIMO	Multiple input multiple output
NIC	Network interface card
Bi-LSTM	Bidirectional long short-term memory
GRU	Gated recurrent units

Tx	Transmitter
Rx	Receiver

## 1 Introduction

Human Activity Recognition (HAR) is a modern technology with various domains, including smart cities, health care, security surveillance, virtual reality, gaming, and location-based services [6]. Wearable sensor-based, and camera-based methods are the conventional approaches for HAR. Although these approaches are promising and widely used, they have limitations that make them unsuitable for all application scenarios. Wearable sensor-based approaches require users to wear sensors on their bodies, which can be uncomfortable or inconvenient in some situations. On the other hand, camera-based methods require cameras to be installed in the environment where the activity is taking place, which can be intrusive and raise privacy concerns.

In recent years, researchers have focused on various applications of WiFi sensing, such as remote control in smart homes [14, 16], localization [18, 24], monitoring driving

✉ Yan Chiew Wong  
ycwong@utem.edu.my

<sup>1</sup> Centre for Telecommunication Research and Innovation (CeTRI), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer (FKEKK), Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia

<sup>2</sup> Faculty of Engineering and Technology, Multimedia University, Selangor, Malaysia

conditions [32], gesture recognition [1, 15], activity recognition [5, 30, 31] and other human–computer interactions and medical applications [24]. WiFi signals can be used as a short-range passive radar by measuring their interaction with movement and the environment [14]. Moreover, the systems track locations and movements by analyzing how signals are reflected and deflected environment [29, 39].

Using WiFi for HAR with location independence is a promising area that has the potential to enhance the performance and applicability of activity recognition systems. This technology provides a robust and reliable approach that accurately recognizes human activities regardless of the sensor’s condition, making it an attractive solution for various real-world scenarios. Mainly, the CSI component comprises several parameters, including amplitude, phase, and delay information, which is used to determine the quality of the wireless link. To use CSI for HAR, researchers use machine learning algorithms to analyze changes in the CSI as a person moves through the wireless environment [13, 14]. By detecting changes in the CSI’s amplitude, phase, and delay information, it is possible to accurately recognize human activities such as walking, running, and sitting. Location dependency makes it difficult to generalize the model to different locations and situations and restricts the technology’s ability to be transferable across multiple sites, which is crucial for commercialization [5, 10, 40].

Location-independent sensing refers to the ability of a technology to work in different environments, sites, or locations without any changes in its configuration or performance [5, 27]. Existing Wi-Fi-based HAR approaches rely heavily on training data collected from specific locations or environments. This dependence on location limits the generalizability and scalability of the models. A robust HAR system should be able to accurately classify activities in different environments and adapt seamlessly to new environments without retraining [20]. Overcoming location dependency is crucial for deploying Wi-Fi-based HAR systems in diverse real-world settings. In addition, accurately classifying static activities, such as sitting, standing, or lying down, poses a substantial challenge in HAR. These activities share similar features and can be difficult to distinguish solely based on Wi-Fi signals. Existing models need help categorizing static activities, leading to lower accuracy in these categories. Improving the classification of static activities is essential for achieving comprehensive and reliable activity recognition [41].

To overcome this limitation, we propose an approach that leverages activity-adapted learning to enable feature transfer between different locations and users based on a RNN-LSTM structure. This approach allows accurate recognition of activities based on location, balances performance with reducing the need for a large amount of training data and allows learning from the user’s interactions. This framework

introduces a systematic approach for accurately classifying dynamic and static activities. It leverages logical sequence classifiers and LSTM-based feature extraction to enhance activity recognition performance. The main contributions of this work can be summarized as follows:

- To present a logical approach used to reconstruct the CSI data, which allows static activities mapping based on a learning algorithm that adapts to new locations. The proposed method employs a coarse-to-fine logical strategy generally applicable to various activity recognition systems.
- To design a location-independent real-time monitoring system that utilizes deep learning technology based on HAR. The proposed model incorporates LSTM networks for feature extraction from WiFi signals in our framework. The utilization of LSTM networks enables the capturing of long-term dependencies and temporal patterns within the sequential data. This advanced feature extraction technique enhances activity recognition accuracy by effectively modeling the complex relationships and dependencies in Wi-Fi-based activity data.
- To validate the proposed framework’s effectiveness and robustness by conducting extensive evaluations and comparisons with existing methods, including other RNN-based approaches. Through this comparative analysis, we demonstrate the superiority and advantages of our framework in terms of activity recognition performance and address the limitations of previous methods. The evaluation provides empirical evidence of the efficacy of our approach and its potential to overcome the challenges in Wi-Fi-based HAR.

By addressing the challenges related to location dependency, improving the classification accuracy of static activities, and leveraging advanced LSTM-based feature extraction, our work significantly contributes to the advancement of Wi-Fi-based HAR. Our proposed framework offers a more accurate and robust solution for activity recognition, thereby facilitating the deployment of Wi-Fi-based HAR systems in various real-world scenarios. To provide a concrete illustration of the motivations and practical implications, let’s consider a smart home equipped with WiFi sensors that monitor the activities of its residents. In contrast to existing sensor-based approaches, which require individuals to train models in every new location with every new activity, our proposed Wi-Fi-based HAR system eliminates the need for intrusive and inconvenient training efforts. The proposed model enables the model to generate fingerprinted dataset based on the new location by detecting dynamic activities. This non-intrusive nature improves user comfort and enhances the overall user experience regarding activity monitoring and behavior analysis within a smart home environment.

The remainder of this paper is structured as follows: Sect. 2 discusses the theoretical background and related works, followed by the problem analysis. Section 4 presents the methodology and experimental setup. Section 5 describes the results and discussion. Finally, limitations and future works are highlighted, and concludes the paper.

## 2 Related Works

WiFi-based sensing is location-dependent compared to sensor-based methods due to its sensitivity to the user's orientation and environmental changes, posing a challenge to the technology's transferability across various locations. However, research continues investigating new methods to overcome these limitations. These methods include developing models independent of the user's location and orientation and mapping the relationship between WiFi measurements and human actions or activities [7]. Practical approaches for addressing location dependency have been proposed in [9, 22], which applied transform theories to practical applications. The recognition algorithm learns location and person independent features from different perspectives of CSI data. The state machine learns temporal dependency information from history classification results. According to the superposition of multipath, the received signal and its effect on the wireless channel for the same activity are greatly modified in different stages, and activity detection is tied to the trained location [17, 20, 25, 35].

Yang et al. developed FALAR, which leverages class-estimated basis space singular value decomposition to eliminate location information from the CSI data associated with static paths [35]. The system was tested on samples of five activity categories collected from eight locations, where four of the locations were used for training. The results showed that FALAR achieved a gesture recognition accuracy of 90.6% for all eight locations. However, the system requires using the new OpenWrt firmware to obtain fine-grained CSI data from all 114 subcarriers. Lu et al. proposed WiHand to enhance gesture recognition in dynamic settings by separating background signals from gesture signals using low-rank and sparse decomposition [20]. Their tests revealed an average testing accuracy of 93% for untrained locations. However, the system relies on high signal transmission rates, which may lead to data packet loss.

Zhang et al. conducted additional research in this field, introducing Widar3.0, a gesture recognition system that uses body coordinate velocity profile (BVP) signals, and a CNN-GRU network to extract spatial and temporal features for classification [41]. Widar3.0 achieves an 85.3% average accuracy for recognizing gesture samples at the fifth location but requires at least three receivers and predefined zones for BVP data collection. Although Widar3.0 improves recognition

performance by separating activity signals from background information, it is limited by specific hardware or deployment targets. Ding et al. introduced WiLiMetaSensing, a method that utilizes a CNN and LSTM network to extract location-independent features for activity recognition [5]. Samples from source locations are used for meta-learning, and only a few samples from target locations are required for training. When four locations are used for training and 24 locations for testing, WiLiMetaSensing achieves an accuracy of 91.11% in one-shot learning. This approach decreases the number of target location samples required for activity recognition but still requires a small number.

Another approach to improving location independence beyond model-based algorithms is array antennas and multiple systems. In [8] and [33], antennas were utilized to improve location dependency by focusing detection on the person's precision. However, this approach does not fully operate location-dependently. In [4], the employment of UWB 5G transmission was used to improve the detection of multi-person estimation. Their work aims to achieve a broader detection bandwidth with more reflected signals over a large band of frequencies. Table 1 summarizes the relevant compositions on location independent HAR and provides information about the type of algorithm used, the type of signal (e.g., CSI), and the equipment required for the data collection to give a general overview of the complexity and feasibility of the method.

## 3 Primarily and Problem Analysis

### 3.1 CSI Mathematical Analysis

We leverage the ubiquity of CSI as the primary means for capturing activity data. CSI is a crucial indicator of the channel link states in MIMO systems. CSI provides a high level of sensitivity to the variations of the channel link, making it superior to other signals due to its fine-grained nature and relatively small size. The mathematical representation of CSI is shown in Equation.

$$y = Hx + n \quad (1)$$

where  $y$  represents the vectors of the transmitter and receiver, The transmission equation represents  $H$  as a complex matrix of CSI values and  $n$  as the channel noise [19]. MIMO enables multiple channels to increase transmission rate by creating  $H$  matrix of connection links, represented as Eq. 2:

$$H_i = \begin{bmatrix} h_i^{11} & h_i^{12} & \dots & h_i^{1N_T} \\ h_i^{21} & h_i^{22} & \dots & h_i^{2N_T} \\ \vdots & \vdots & \vdots & \vdots \\ h_i^{N_R 1} & h_i^{N_R 2} & \dots & h_i^{N_R N_T} \end{bmatrix} \quad (2)$$

**Table 1** Benchmarking HAR with location independency utilizing CSI

References	Method	Classifier	Tool	Freq GHz	Activity	Accuracy %	Performance and limitations
[5]	Location variations with the limited dataset	AdaBoost	Intel 5300 NIC	5	Gesture	91	Few dynamic activities, requires a small number of target location samples. limited to multiple antennas, limited to dynamic activities
[22]	Multi-layers algorithm prediction	Reinforcement learning	Intel 5300 NIC	–	Activities	97	Requires four receivers, RL requires massive training effort with huge data at each receiver
[20]	Separates gesture signal from background information	LRSD decomposition,	Atheros AR9380	–	Gesture	93	Needs the high transmission rates of signals (such as 2500 packets/s), which will result in the loss of data packets., limited to dynamic activities
[25]	Doppler frequency shift	CNN, CNN-LSTM	Intel 5300 NIC	5.32	Gesture	Average of 94	Dependent to dynamic activities, which means, static activities are unrecognized Using multiple Rx Aps and tested in the same location
[41]	Using body velocity pattern for predicting activities	CNN	Intel 5300 NIC	5.82	Gesture	87	Limited to specific hardware equipment. Requires multiple Aps. Different people have different body velocity and speed performing actions. Only tested by dynamic activities

The CSI estimates the magnitudes and phases  $h_i^{N_R N_T}$  of the  $i$ th subcarrier for the link between the receiver antenna and the transmitter antenna [28]. Hence, the CSI entry corresponds to the channel frequency response, as Eq. 3 indicates.

$$h(f) = \sum_{l=1}^N \alpha_l \exp^{-j2\pi f \tau_l} \quad (3)$$

where  $N$  represents the summation of total multipaths of the subcarriers,  $\alpha_l$  represent the attentions and  $\tau_l$  is the propagation delay of signal through path  $l$ . The WIFI CSI ratio illustrates how surrounding objects affect, weaken, and scatter OFDM signals during transmission [14, 38]. On the one hand, the uncertainty in the power amplifier of the RF chain regularly leads to impulsions and burst noise in the amplitude of the CSI. Additionally, the disparity in the frequency band between transceivers causes a time-varying phase to offset in each CSI sample, which quickly supports this perspective and disrupts the phase variation caused by human motion and mathematically represented in Eq. 4 as:

$$H(f, t) = \delta(t) e^{-j\phi(t)} \sum_{l=1}^L A_l(t) e^{-j2\pi \frac{d_l(t)}{\lambda}} \quad (4)$$

whereby  $\delta(t)$  represents the intensity of impulsive noise and  $\phi(t)$  represents the time-varying phase offset.  $L$  represents the total number of propagating routes,  $\lambda$  represents the wavelengths, and  $A_l(t)$  and  $d_l(t)$  represents the signal's attenuation and the  $L$  pathway's length, respectively. Overall, the environment significantly impacts the wireless signals used to capture the CSI measurements. Furthermore, the accuracy of human activity recognition is also influenced by the positioning and orientation of the person being monitored.

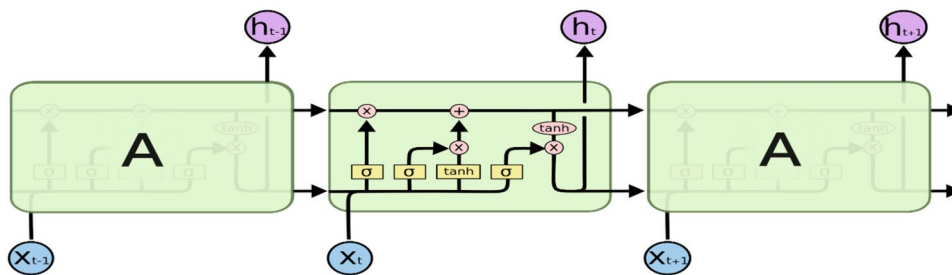
### 3.2 LSTM Networks

LSTM networks are particularly well-suited for this task due to their ability to capture and analyze long-term dependencies within sequential data effectively. This characteristic is crucial for accurately identifying patterns and trends within the analyzed signals. Within the LSTM architecture, the facilitation of information removal or addition to the cell state is proficiently governed by specialized components called gates shown in Fig. 1. These gates act as discretionary conduits for the passage of information. Structurally, they are comprised of a sigmoid neural network layer and a pointwise multiplication operation, operating in tandem to exert meticulous control over the information flow [11].

The input set of the CSI vector is defined as  $\mathbf{x} = x_0, h_1, h_2 \dots h_t$ , the output set is  $\mathbf{y} = y_0, y_1, y_2 \dots y_t$ . And the hidden layers  $ash = \{h_1, h_2, \dots h_t\}$ . In the initial



**Fig. 1** Interacting layers within an LSTM module [11]



stage of LSTM model, a crucial determination is made regarding the selective discarding of information from the cell state. This pivotal decision employs a sigmoid layer called the “forget gate layer.” By examining the previous hidden state  $h_{t-1}$  and the current input  $X$ , this layer generates an output value ranging between 0 and 1 for each component within the initial cell state  $C_{t-1}$ .

In contrast to the input aggregation and processing mechanisms employed by RNNs, LSTM networks exhibit a more suitable architecture for prolonged data recognition as an input gate. Incorporating a forget gate allows for comparing the internal memory and new incoming data, facilitating selective overwrite. This dynamic process enables the smooth propagation of gradients across sequential steps. The LSTM comprises an input gate, forget gate, output gate, and memory cell to effectively manipulate the data to be forgotten, recognized, and retained. The gating technique, encompassing both the activation function (sigmoid function) and element-wise multiplication, is selected to govern the flow of pertinent data. The resulting output value, confined within the  $[0, 1]$  range, facilitates subsequent multiplication and regulates the data flow. The relevant initialization gates are assigned values close to or equal to 1, thereby mitigating any detrimental impact on initial training stages.

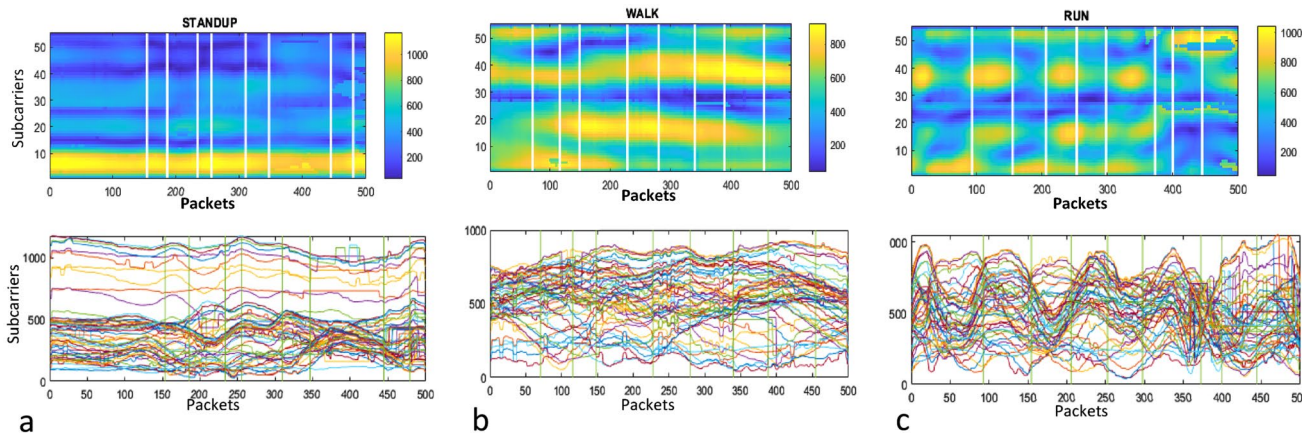
**3.3 Correlation of Dynamic Activities and CSI with Environment**

The analysis of activities, as depicted in Fig. 2, indicates that dynamic activities confirm exceptional diversity in their packet transmission. The diversity in packet transmission facilitates precise discrimination of various activities. Moreover, the dynamic activities exhibit variations in both amplitude and phase, which consequently manipulate the velocity of movement.

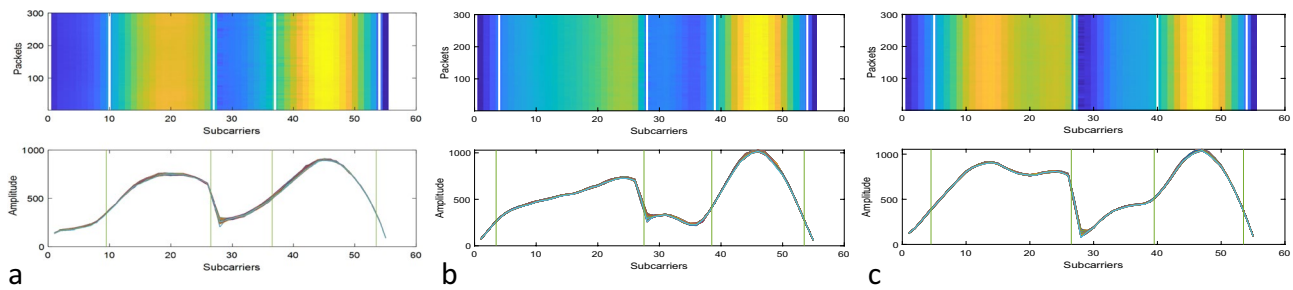
In contrast, stationary activities, including empty, sitting, and standing perspectives, display a lower degree of variability in packet transmission, making it more challenging to distinguish them from one another. This underscores the need to account for both the activity and the location when analyzing CSI data from wireless networks. As illustrated in Fig. 3, subcarrier variations are observed for distinct stationary postures, such as empty, sitting, and standing, in a location where the frequency remains unchanged.

**3.4 Environmental Effects of WiFi Sensing**

Utilizing WiFi for human tracking and localization presents many challenges that necessitate careful consideration to achieve accurate and reliable performance. Firstly, multipath propagation occurs due to the intricate interplay



**Fig. 2** Variation in amplitude in dynamic activities **a** standup, **b** walk, and **c** run



**Fig. 3** The amplitude and phase changes among static activities such as **a** empty, **b** standing, and **c** sitting

of reflections, diffractions, and scattering within indoor environments, as indicated in Fig. 4. As a result, WiFi signals traverse multiple paths and exhibit disparities in signal strengths, rendering the accurate estimation of an individual's location based solely on WiFi measurements a formidable task.

Moreover, NLOS conditions emerge when obstacles or physical barriers obstruct the direct path between WiFi access points and the tracked individual. WiFi signals experience notable attenuation and distortions in such scenarios, culminating in inaccurate localization estimates. Additionally, the vulnerability of WiFi signals to interference and noise from diverse sources, including other WiFi devices, electronic appliances, and environmental factors, poses significant challenges. Such interference and noise detrimentally affect the quality of the received signal, consequently leading to localization errors. Another challenge is the complex and unpredictable propagation of wireless signals within indoor environments is affected by severe signal attenuation, reflection, and multipath effects. Equation 5 represents the Power propagated in space with varied gains between transmitter and receiver. By determining the environmental aspect, it is possible to analyze the environmental effects of the signal to eliminate the dependency on the environment. Starting from Power received Eq. 5.

$$P_r = \frac{P_t * G_t * G_r * \lambda^2 * F}{(4\pi * R)^2} \quad (5)$$

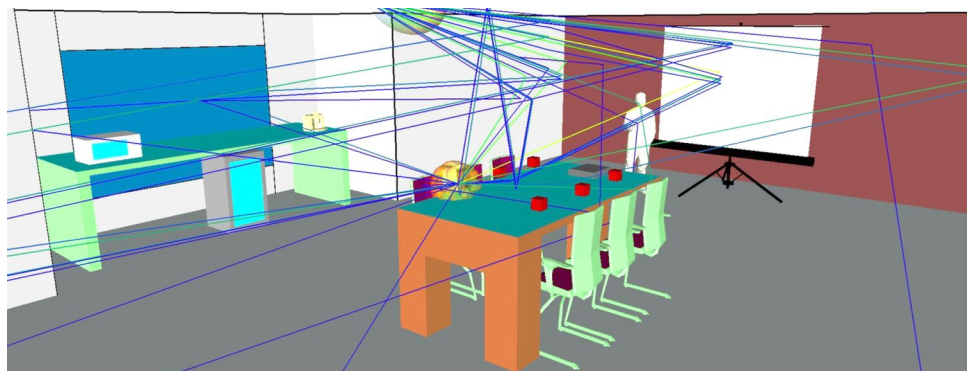
$P_r$  represents the received power,  $P_t$  denotes the transmitted power,  $G_t$ ,  $G_r$  refer to the transmitted and received gains,  $\lambda$  is the signal wavelength and  $F$  is the propagation factor. The  $R$  represents the propagation range. Additionally, one of the significant challenges in HAR is the variability in how people perform activities. People may perform the same activity at different speeds, orientations, and methods, which leads to signal variations and make it difficult to predict the type of activity being performed accurately.

## 4 Methodology

### 4.1 System Overview

The process begins with data collection using RPi supported by Nexmon firmware and TCPDUMP, as depicted in Fig. 5. The raw CSI data is then subjected to MATLAB preprocessing and filtering stages. These stages aim to enhance the quality and relevance of the collected data, ensuring its suitability for subsequent analysis. The behavior recognition system's critical component is utilizing LSTM networks for feature extraction from the signal data. The architecture described herein presents a sequential model that utilizes LSTM layers for classification, as depicted in Fig. 5. The

**Fig. 4** Multi-path propagation of WiFi



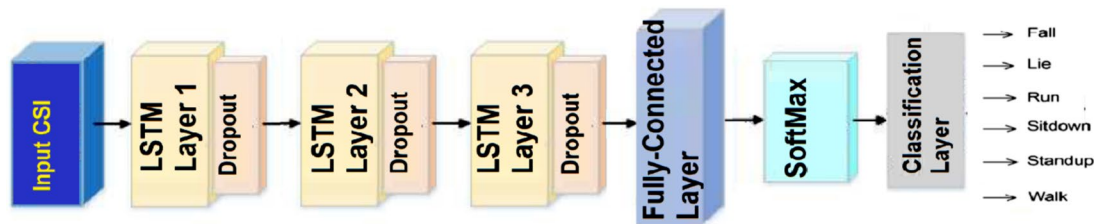


Fig. 5 Schematic of LSTM architecture layers

workflow commences with an initial input layer designed to receive data input representing the length of the sequence. Subsequently, this sequential data is passed through an LSTM layer, a specialized recurrent neural network layer designed to handle sequential information. Notably, the LSTM layer incorporates memory cells that enable the retention of information over time, thus effectively capturing long-term dependencies within the data. A dropout layer is comprised after the LSTM layer to introduce regularization and mitigate overfitting. Dropout randomly deactivates specific input units during training, thereby promoting the independence of neuron learning.

The subsequent layer, “Fully Connected LSTM Layer 2,” applies fully connected operations to the output from the preceding LSTM layer. This fully connected layer facilitates detecting intricate relationships and complex patterns within the data by establishing connections between all neurons. Moreover, the model features an output layer comprising two fully connected neurons, serving as a classifier to categorize the input data into one of two classes. An LSTM layer labeled as “LSTM 3” is also present, presumably aimed at capturing further temporal dependencies and information. Dropout regularization is applied again in the subsequent layer labeled “Dropout 3.” Finally, a fully connected layer is employed before a softmax layer, which outputs probability distributions across the seven possible classes. Ultimately, the architecture culminates with the output layer, utilizing a classifier to assign the input data to one of the seven predefined output classes.

In the subsequent stage of model training and hyperparameter tuning, the raw training data is further divided into 80% for training purposes and 20% for validation to evaluate the trained model. Five LSTM-based models are evaluated using the validation data, and the hyperparameters of the trained models are subsequently optimized using the optimization approach. Finally, the hyperparameter-tuned models are assessed against the test results, and their respective performance in activity recognition is compared. The proposed scheme for location-independent human activity recognition comprises two main phases: offline training and online testing, as depicted in Fig. 6. The system’s workflow encompasses four key components:

data collection, data preprocessing, feature representation, and model training/testing. During the data collection, raw CSI measurements are gathered to capture the environmental variations. Data preprocessing involves calculating the amplitude using the raw CSI and employing median and outlier hamper filters to eliminate noise. The collected data is then partitioned into time  $\times$  subcarrier size samples, which indicate the number of frames corresponding to an activity multiplied by the number of subcarriers.

The data samples are then mapped to a high-dimensional embedding space using LSTM for feature representation. Concatenating models is a technique that combines multiple models’ output to achieve higher accuracy in classification tasks shown in Fig. 6. In this context, the presented approach employs concatenating models to classify groups of activities into smaller categories. The initial model is the fingerprinting model, which utilizes amplitude subcarriers of the CSI to confirm whether a given location is currently empty or has been occupied by a user. The user would be required to capture several samples to establish a baseline characterization of the empty environment at that site. Furthermore, the next classifier, model 2, differentiates between dynamic and static activities. The third classifier (model 3) classifies dynamic activities into those with shifting movement (e.g., walking, running) and those without moving movement (e.g., sitting down, standing up, falling). Another concatenated model is trained to classify walking and running activities using model 4 for dynamic activities with movement. Additionally, logical and sequential classifications of activities with similar features, such as sit-down and stand-up, and falls, are utilized to classify these activities based on human logical concepts.

The offline model concatenates the trained models with significant blocks. Model 5 generates a new dataset for static activities from dynamic activities and assembles a trained model based on current environmental parameters. Overall, the proposed system offers a robust and effective solution for location-independent human activity recognition, with the potential to be applied in various real-world scenarios.

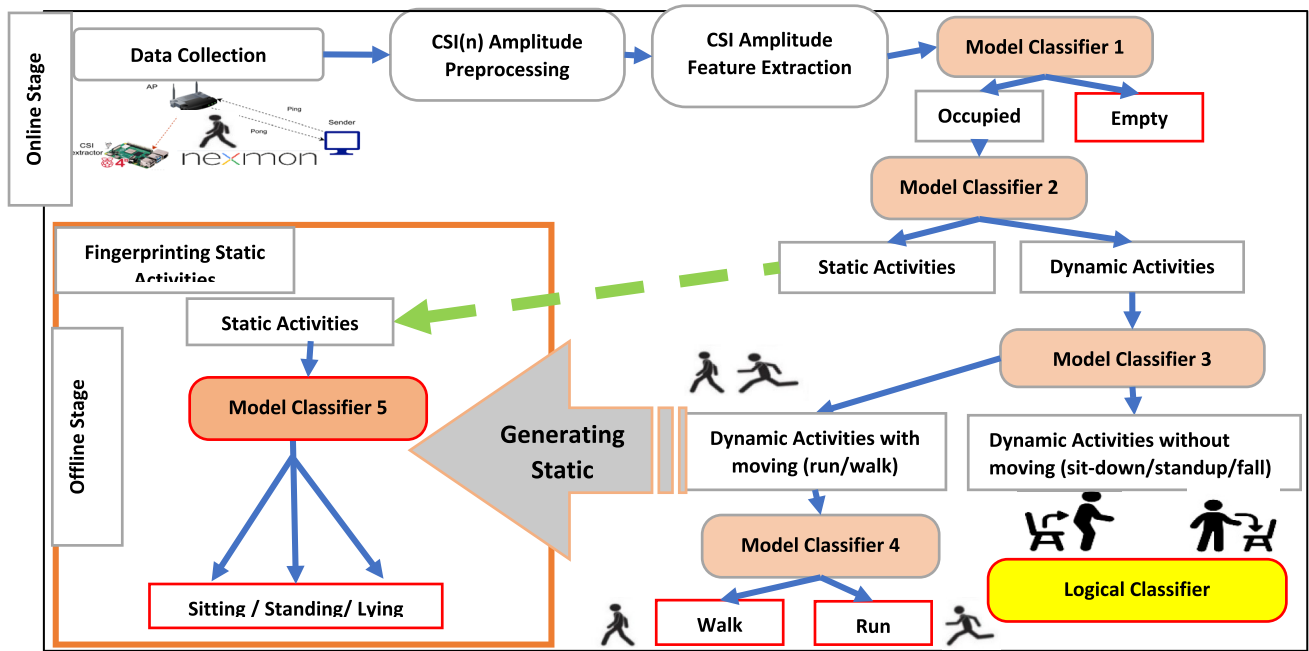


Fig. 6 System flowchart

## 4.2 Preprocessing

A series of data-cleaning techniques were employed to optimize the accuracy and reliability of the combined CSI data. We first utilized a median filter approach to remove potential outliers in the data, substituting anomalous values with the median value derived from the surrounding data points. Despite this initial filtering step, some outliers persisted in the dataset. To moderate the impact of these remaining outliers, we employed Hampel fill outlier filters that replaced any such values with the initial non-outlier value in the dataset. Furthermore, a moving median filter technique was applied to refine the data and reduce any remaining noise, which involved calculating the median value over data points. This filtering approach effectively minimized minor fluctuations or inconsistencies in the data. The standard deviation measures how much the amplitude deviates from the mean value during data acquisition, making it helpful in identifying amplitude outliers. The denoising process starts by calculating the mean value of the  $i$ th subcarrier of  $k$ th data packet according to Eq. 6.

$$\overline{CSI\text{Amp}}^i = \frac{1}{N} \sum_{k=1}^N CSI\text{Amp}_k^i \quad (6)$$

where  $N$  is the number of samples, and  $I \in [1, 2, \dots, 56]$  is the subcarrier index. Then we calculate the standard deviation of the  $i$ th subcarrier from Eq. 7.

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{k=1}^N CSI\text{Amp}_{k,k=1}^i \left( CSI\text{Amp}_k^i - \overline{CSI\text{Amp}}^i \right)^2} \quad (7)$$

$i$  is the index of the subcarriers, so we can get the  $V = [\sigma_1, \sigma_2, \dots, \sigma_{55}, \sigma_{56}]$  which is a variance matrix of the 56 subcarriers. Assuming that the data packet to be filtered is  $k$ , the CSI amplitude values are  $|CSI\text{Amp}_{k-1}^i|$  and  $|CSI\text{Amp}_{k+1}^i|$  for each adjacent data packet  $k-1$  and  $k+1$ , respectively. According to Eq. 8, the filtered amplitude  $|Amp|_{\text{filter}}$  is calculated by averaging the three amplitude data values.

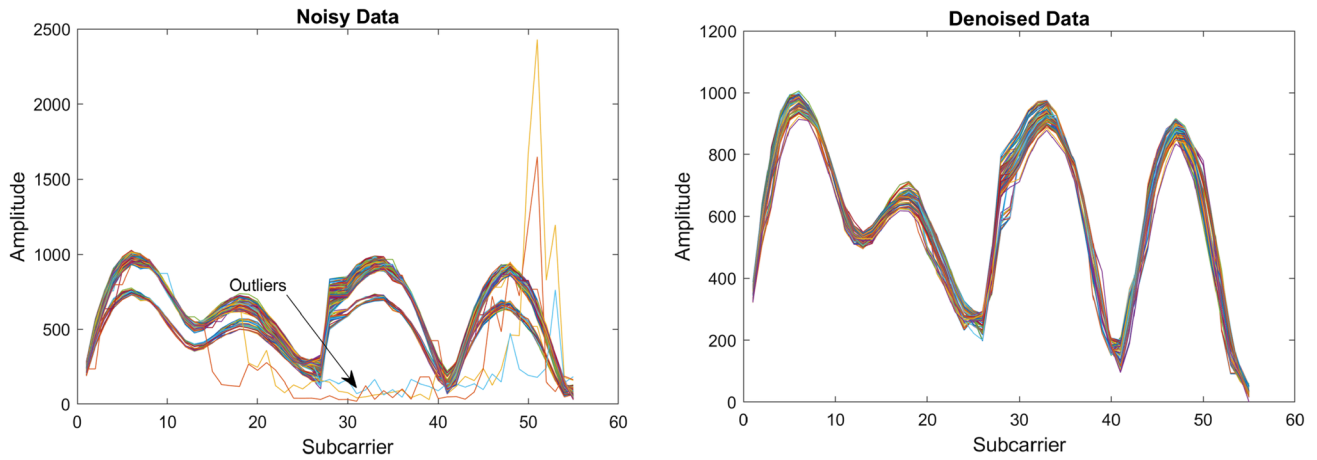
$$|CSI\text{Amp}|_{\text{filter}}^i = \frac{1}{3} (|CSI\text{Amp}|_{k-1}^i + |CSI\text{Amp}|_k^i + |CSI\text{Amp}|_{k+1}^i) \quad (8)$$

For the processed amplitude  $CSI\text{Amp}_{\text{filter}}$ , the covariance matrix  $\text{Cov}(CSI\text{Amp}_{\text{filter}}, CSI\text{Amp})$  of  $CSI\text{Amp}_{\text{filter}}$  and  $CSI\text{Amp}^i$  is calculated. After processing the CSI amplitude values shown in Fig. 7, the filtered CSI amplitude values are smoother, with the redundancy caused by various factors effectively removed and the abnormal values caused by environmental factors filtered out.

## 4.3 Feature Extraction

LSTM is an advanced, recurrent neural network that excels at capturing long-term dependencies in sequential data. In the CSI context, LSTM enables feature extraction by leveraging





**Fig. 7** Denoising process and removing outliers of CSI amplitude

its memory retention, non-linear mapping, contextual understanding, and ability to handle variable-length sequences. It remembers relevant information, captures complex relationships, understands temporal dynamics, and adapts to varying sequence lengths. Compared to traditional methods and simpler models, LSTM's strength lies in its capacity to uncover patterns, extract meaningful representations, and exploit temporal dependencies, making it well-suited for feature extraction in CSI analysis [26].

To maintain the collected data's integrity and prevent the loss of crucial signal features, we opted against utilizing principal component analysis (PCA) or linear discriminant analysis (LDA) that reduces data dimensionality. Instead, we have used the complete set of collected data, organizing it systematically and coherently based on the amplitude of the CSI. Algorithm 1 outlines the denoising procedure in a systematic manner and denoted for the expression of feature extraction of CSI feature matrices.

---

**Algorithm 1:** CSI Preprocessing, outlier removal and feature extraction.

---

**Input:** Raw CSI<sub>Amp</sub> data ← local median CSI<sub>Amp</sub><sup>l</sup>

**Output:** Denoised CSI Data

- 1: CSI<sub>Amp</sub><sup>Hi</sup> ← CSI<sub>raw</sub> - CSI pilot - CSI null subcarriers to remove pilot and null subcarriers
  - 2: Comparing the current CSI<sub>Amp</sub>(i) with  $\sigma_i$  using Equation 6
  - 3: **if**  $CSI_{Amp}_k^i - \overline{CSI_{Amp}^l} > k\sigma \times \sigma_i$
  - 4:      $CSI_{Amp}^l = CSI_{Amp}(i)$
  - 5: **end if**
  - 6:  $CSI_{Amp}(i) \leftarrow \overline{CSI_{Amp}^l}$  outlier Hampel filter
  - 7:  $CSI_{Amp\_hampel\_Denoised}(i) \leftarrow CSI_{Amp}(i)$  median filter
  - 8: CSI<sub>out</sub> [56 x 200] ← Feature extraction from CSI Amplitude
-

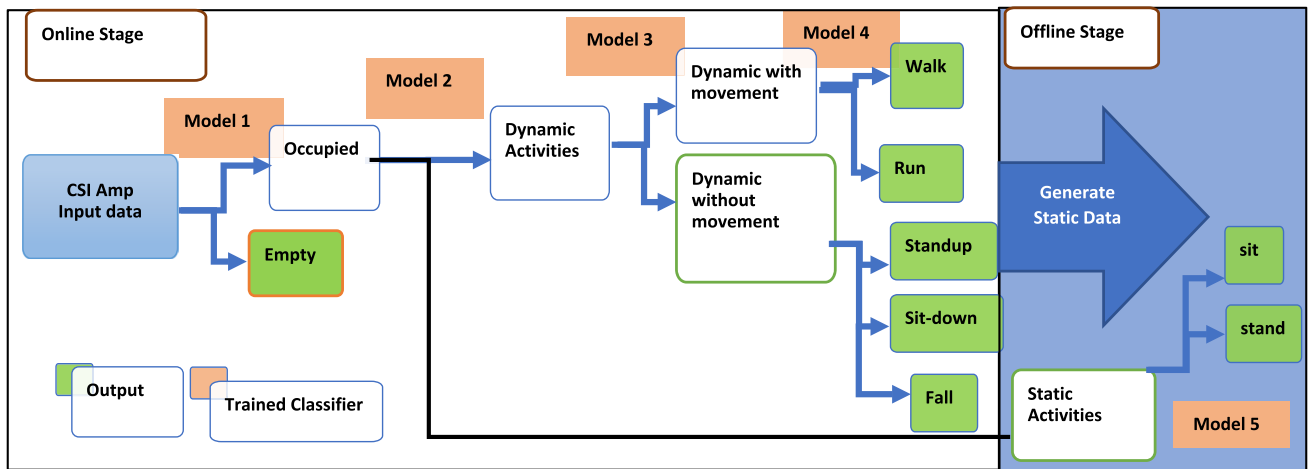


Fig. 8 Diagrammatic representation of training classifier concatenation via model concatenation

## 4.4 Model Classifier

### 4.4.1 Online Stage

The online stage pertains to the current state of a model undergoing training by feeding it with a dataset and adjusting its parameters, enabling it to recognize patterns and make predictions on new data. In this work, the online stage is used to train models that classify various activities according to their characteristics, beginning with general activities such as empty locations and progressing to more specific ones that involve dynamics and static movements. The training process concatenates multiple models in a sequence, with the output of one model used as the input for the next model, resulting in a connected model that recognizes complex patterns and makes accurate predictions. Figure 8 illustrates the concatenation process, where each model is depicted as a node or block in the diagram.

**4.4.1.1 Trained Model 1—Empty Location Classifier** CSI offers valuable insights into the occupancy status of a location by analyzing the strength and quality of wireless signals. After preprocessing the data and using the sub-carrier amplitude, the LSTM model was trained to classify locations as empty. The underlying assumption is that

any obstruction or blocking of wireless signals due to the presence of occupants in a site will cause changes in the amplitude of the CSI, enabling the model to detect the empty status of the location accurately.

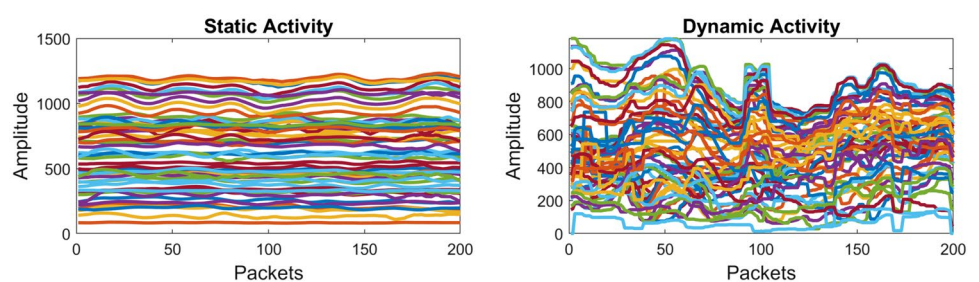
**4.4.1.2 Trained Model 2—Static/Dynamic Classifier** The initial dissection of concatenated models involves the Model 2 classifier, which aims to categorize activities as either static or dynamic. Although the activities within each group are unspecified, this model leverages frequency patterns to classify them effectively into two overarching groups with exceptional accuracy. It should be noted that this classifier is considered location-independent, as the frequency modulation feature operates independently of location, as shown in Fig. 9.

To classify dynamic activities, we utilize moving variance segmentation MVS of packet changes, which is more sensitive to subtle body movements because it uses a sliding window to compute the sum of squared signals. Specifically, we calculate MVS within a sliding window using the following procedure in Eq. 9.

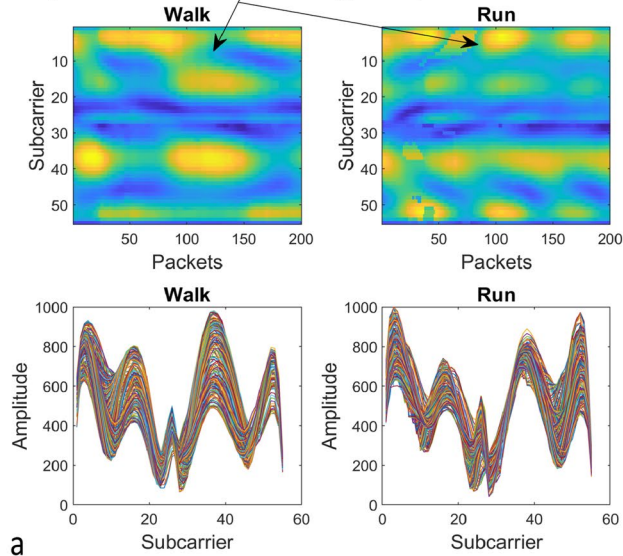
$$CSI_{mvs(t)} = \sum_{i=1}^n \left[ \frac{1}{L-1} \sum_{j=1}^L |CSI_{j \in L} - \mu|^2 \right], \quad (9)$$

$$\mu = \frac{1}{L} \sum_{j=1}^L CSI_j.$$

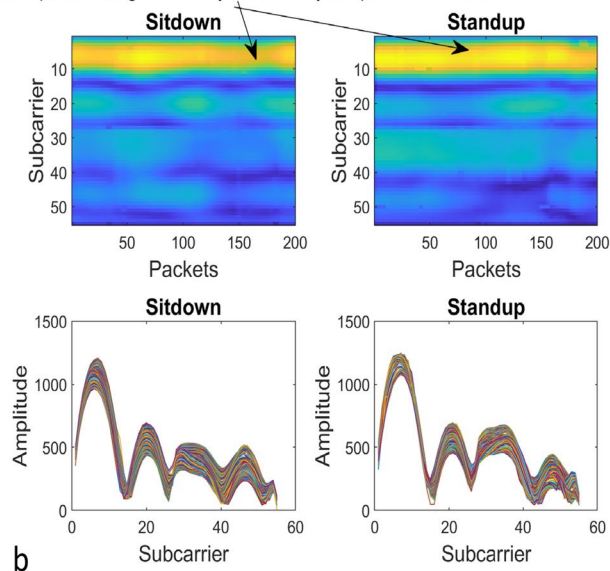
Fig. 9 Variation features of dynamic and static activities



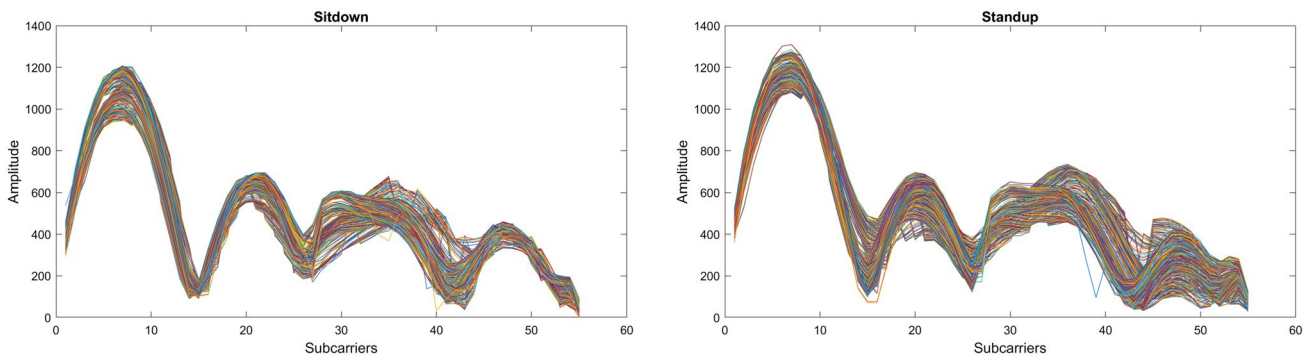
CSI Amplitude and Phase Varies due to shifting person's position.



CSI Amplitude changes due to dynamic activity, but phase is constant.



**Fig. 10** Classifying dynamic activities into dynamic with/without movement based on the CSI amplitude **a** with movement (walk and run) and **b** without movement (sit-down and standup)

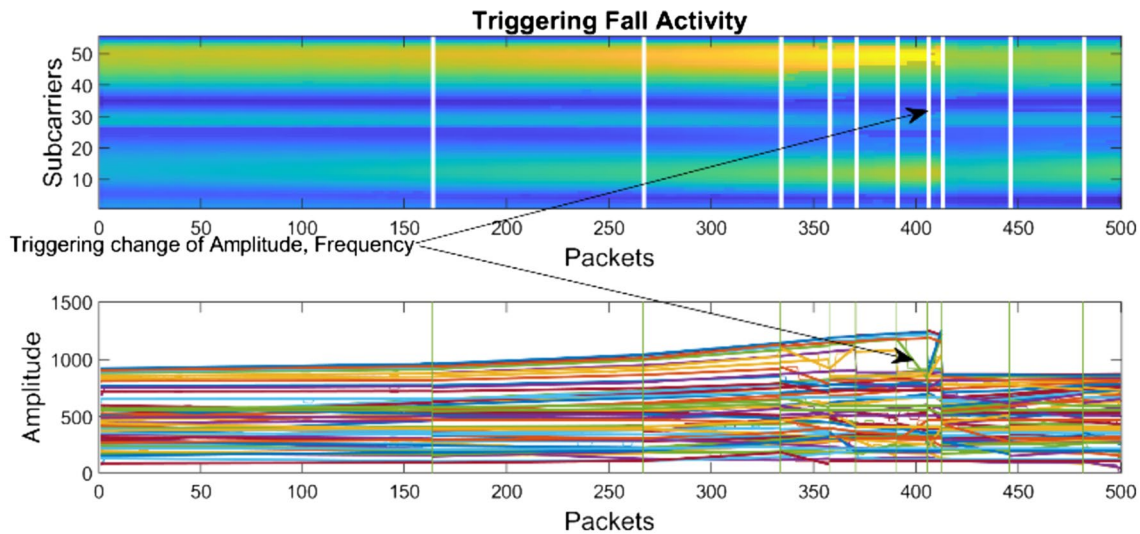


**Fig. 11** Similarities between sit-down and standup, which makes it complex for the trained model to classify

where  $n$  is the length of the captured packet,  $v(t)$  is the moving variance of CSI amplitude at the  $i$ th packets. The moving variance of a CSI stream consisting of  $n$  packets, Eq. (8) defines the variance based on the mean  $\mu$  and packet number ( $j$ ) within a sliding window of length  $L$ . Additionally,  $i$  represents the current position of the packet number within the entire CSI stream. As a result, this observation has stimulated the conception of an adaptable, dynamic threshold in its application across various activities. This threshold is intended to facilitate detecting dynamic activities and consequent data segmentation. Consequently, the subsequent step involves further classifying dynamic activities into more refined subgroups, those with or without shifting positions.

**4.4.1.3 Trained Model 3 (Dynamic Activities Classifier)** Wireless channel variations occur during certain activities, which cause fluctuations in CSI measurements. In order to detect human activities and perform data segmentation, Moving Variance Segmentation (MVS) at Eq. (8) is applied to the moving variance of CSI amplitude with subcarrier. And to achieve that, Model 3 is trained based on the phase shift shown in Fig. 10.

Furthermore, human activities are complex and have similar features, making them difficult to classify accurately. For example, the activities of sitting down and standing up can be very similar regarding the data collected, as shown in Fig. 11. Furthermore, the variation in the sensor data for



**Fig. 12** Triggering changes of activities using change-point function to detect fall activity

these activities can differ between locations, which makes it challenging to generalize activity recognition models across different environments.

To address these challenges, our proposed model uses a logical classify these two similar activities based on the sequence of previous activities.

#### Logical Sequence Classifier

We employed a logical algorithm based on the human activity flow process to improve the classification of dynamic activities. The Algorithm involves a sequential approach to activity classification, considering the relationship between current and preceding older activities to make predictions about the activity in which an individual is currently engaged. By analyzing the logical sequence of actions, our model can accurately classify sit-down and stand-up activities in various locations, regardless of environmental factors. The present Algorithm provides a logical framework for distinguishing between standing up and sitting down at any location, despite their high similarity and inherent classification difficulty. By incorporating changepoint functions, the Algorithm accurately

identifies changes in activity, including falls, by analyzing the sequential patterns of activity depicted in Fig. 12.

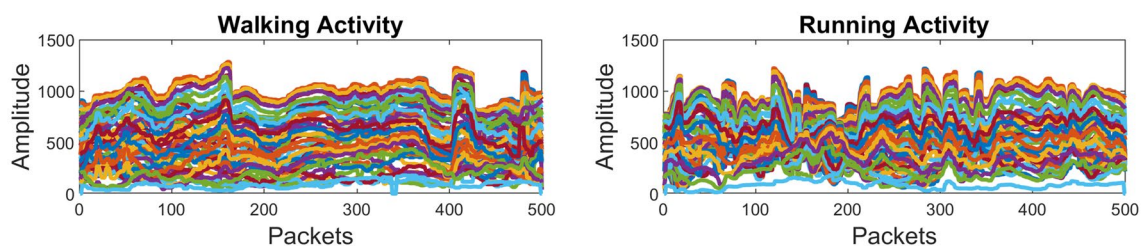
#### 4.4.1.4 Trained Model 4 (Dynamic Moving Activities Classifier)

A narrow classification within the context of the CSI measurements of a subcarrier, denoted as the discrepancy between the observed value and the actual value, is expressed where  $n$  represents the number of samples. The variance is formally defined as the expectation of the squared deviation of a random variable from its mean represented by Eq. 10 [34].

$$S^2 = \frac{1}{n} \sum_{i=1}^n \delta_i^2 = \frac{1}{n} \sum_{i=1}^n (H_i - \hat{H})^2 \quad (10)$$

In contexts, the arithmetic means  $\hat{H}$  is a substitute for the quantity  $\hat{H}.V_i$  defined as  $V_i = H_i - \bar{H}$ .  $\delta_i$  and this relationship expressed mathematically as in Eq. 11.

$$\sum_{i=1}^n \delta_i^2 = \frac{n}{n-1} \sum_{i=1}^n V_i^2 \quad (11)$$



**Fig. 13** Dynamic activities classification based on speed changes



Furthermore, the CSI amplitude variance changes  $S^2$  is estimated using Eq. (12).

$$S^2 = \frac{1}{n} \sum_{i=1}^n \delta_i^2 = \frac{1}{n} \frac{n}{n-1} \sum_{i=1}^n V_i^2 = \frac{1}{n-1} \sum_{i=1}^n (A_i - \bar{A})^2 \quad (12)$$

where  $H_i$  represents the amplitude of CSI matrix with sample  $I$ , and  $\bar{H}$  represents the mean center of the sample, besides  $n$  represents the number of samples. We employed this method to estimate the variance of the local range, which enables discrimination between dynamic activity signals based on the speed of variation shown in Fig. 13.

Algorithm 2 explains the systematic methodology for achieving classifications of dynamic activities within the

framework of training models for Models 1–4. In order to streamline this algorithm, we refer to Fig. 5, which illustrates an online stage for generating four trained models. The process begins with Model 1 to check the occupation of the location. If the site is empty, no further analysis is performed. However, if the place is occupied, Model 2 is employed to determine the type of activity, distinguishing between dynamic and static movements. Next, Model 3 comes into play to classify dynamic activities into moving or non-moving activities. Finally, Model 4 extends the classification further, providing more precise categorization based on the speed of variations as described in Eqs. 9, 10, and 12.

---

**Algorithm 2:** Online Training Model Classifiers

---

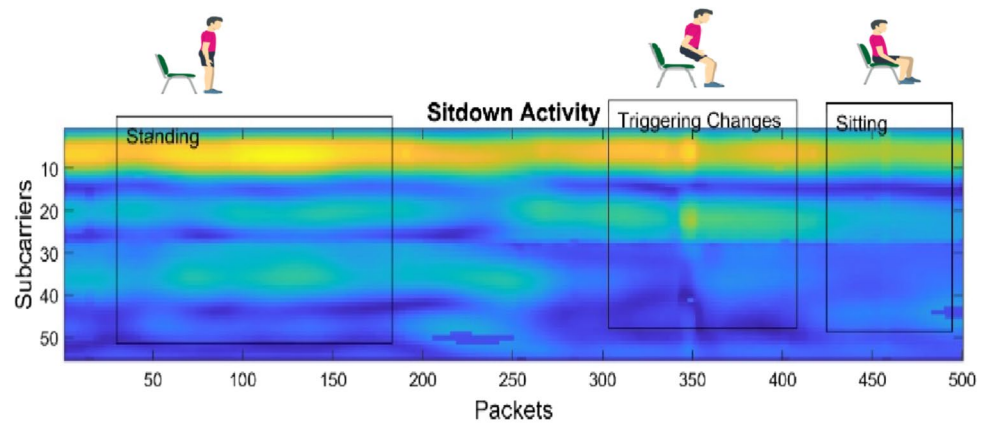
```

1: Input: Denoised CSIout Matrix.
2: Output: Four trained concatenating models.
3: while (true):
4:     [MODEL 1] Check if Location is Empty
5:         Location ← empty, no activity to monitor;
6:     [MODEL 2] Check if Location is Empty
7:     Compute the moving variance of the current window using the MVS technique
8:     if MVS(t) ≠ 0.05 × MVS(t+1) // Classify Static/Dynamic Activities.
9:         If MVS has not changed
10:            reset consecutive MVS → "Static Activity"
11:         else; MVS has changed,
12:            increment  $\sum_{k=0}^n$  nsecutiveMVS and check if it meets the threshold
13:            set the activity classification to MVS → "Dynamic Activity"
14:     end if
15:     [MODEL 3] dynamic activities classifier into (Dynamic with movement /without movement)
16:     Calculate the (MVS) technique for subcarriers phase  $e^{-j\phi(t)}$ 
17:     Compare the current MVS(t) with previous MVS(t-1) amplitude to determine activity type
18:     if  $\delta(t)e^{-j\phi(t)}$  has varied,
19:         increment consecutive Matrix A → Equation (4),
20:         check if it meets the threshold
21:     elseif threshold is reached
22:         set the activity classification to "Dynamic with Movement"
23:     else  $\delta(t)e^{-j\phi(t)}$  has not triggered threshold, reset the activity as Dynamic, but not moving from the
24:         position.
25:     Calculate the changes that occurred in CSI using changepoint.
26:     Set T ← 10 ∈ points at the rms level of the signal changes. //Fall Classifier
27:     [MODEL 4]
28:     To validate the CSI amplitude and the frequency changes of CSI amplitude using  $V_a$ ,  $\delta_i^2$  using Equatin
29:     (9,10)
30:     Estimate variance changes S with threshold representing the speed of movement.
31:     Input CSI amplitude  $H_i$  into the LSTM model for processing.
32:     Use the trained model to obtain the activity based on the S speed of amplitude changes
33:     end
34:     end

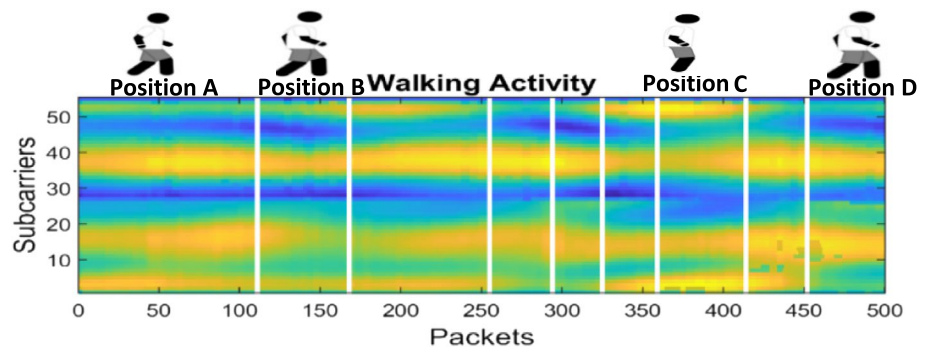
```

---

**Fig. 14** Extracting static activity samples from sit-down activity



**Fig. 15** Generate samples of standing activity in different positions from walking activity



#### 4.4.2 Offline Stage

The trained models are considered robust and location-independent, as they accurately classify dynamic based on their variance of change, as demonstrated in the online stage. To further enhance performance during the offline stage, we utilized the model structure supplemented with two grades: a static activities generator algorithm and self-trained classifier model for these activities. The new stages improves the model's capability to identify and categorize more specific activities.

##### i. Generating Static Dataset

The partitioning of dynamic activity presents an opportunity to redefine it as static transformations between states that are themselves stationary. This notion is exemplified by the sit-down activity, as illustrated in Fig. 14, which comprises three discernible phases: standing, sitting, and the transitional movement connecting the two states. Identifying variations in the transitional direction allows for capturing diverse static postures and locations and acquiring a more comprehensive comprehension of the dynamic activity.

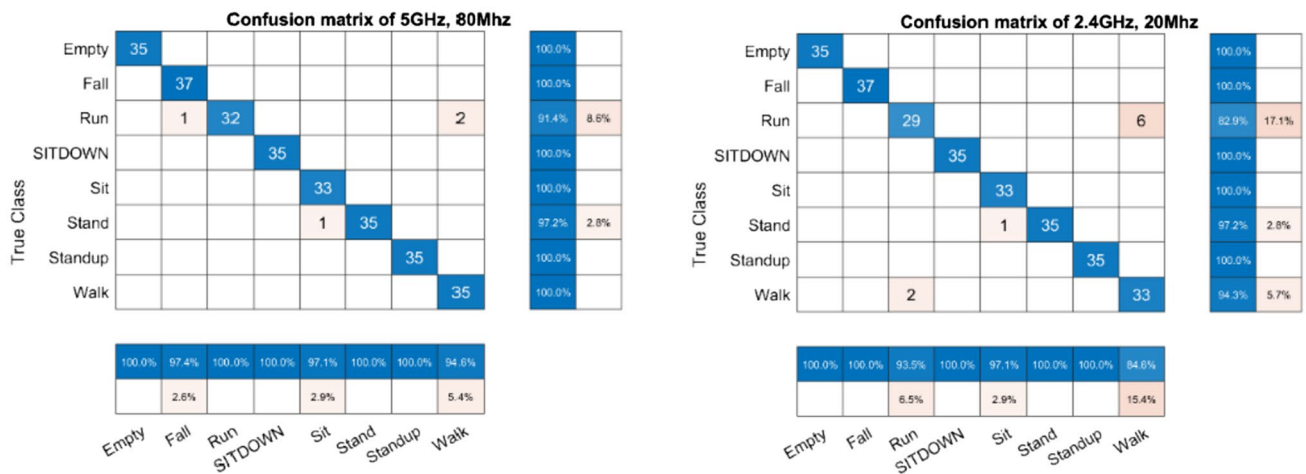
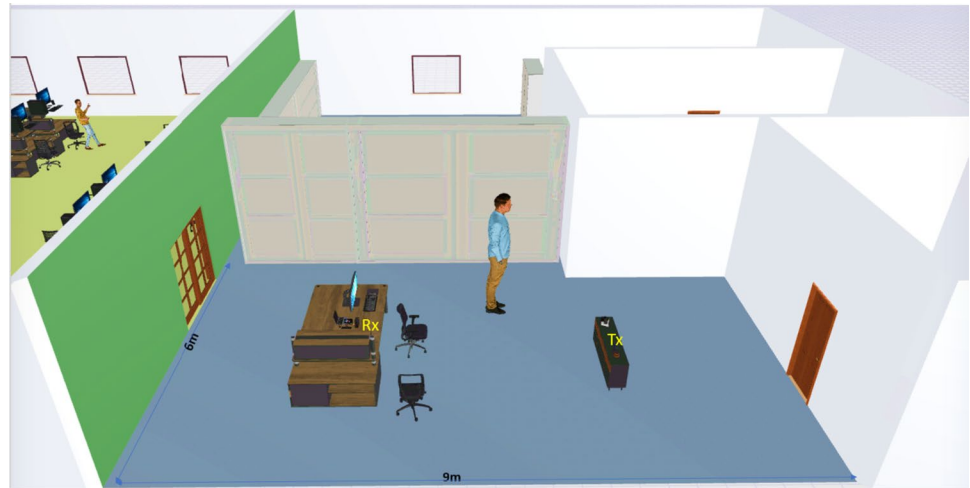
Similarly, the model facilitates the detection of variations in walking activity and subsequently partitions it into constituent standing frames, as illustrated in Fig. 15. This analytical approach engenders a heightened movement level, thereby enabling the monitoring of posture or movement pattern alterations at each position in a path.

This method enables a comprehensive representation of the body's movements and responses to different positions, resulting in a more accurate and effective fingerprinting of new location models.

##### ii. Model 5 (Static Activities Classifier)

The proposed approach in this study involves generating a static dataset using previous model classifiers and a logical classifier that is not dependent on location or orientation. This dataset is then utilized by Model 5, which introduces the advantage of fingerprinting and mapping static activities in a new environment without requiring the model to be explicitly trained for each specific environment. The static activities are classified using an LSTM-trained model. This LSTM model is trained after a sufficient dataset for static activities has been generated to align with the nature of the task. The LSTM model takes input data that describes the static activity under consideration. The input data is then processed through a series of LSTM cells that effectively capture the temporal dependencies within the input data sequence. By employing LSTM cells, the model effectively analyzes and understands the input data's sequential nature, allowing it to capture and utilize the long-term dependencies between different time steps. This capability enables the model to recognize patterns and dependencies within the

**Fig. 16** The layout of training location at lab-hall



**Fig. 17** Confusion matrix analysis of HAR model performance at trained location

input sequence, thereby facilitating the accurate classification of static activities. Algorithm 3 breaks down the

process of generating datasets of static activities and their type using Model 5.

**Algorithm 3:** Static activities classifier.

**Input:** CSI Amplitude concatenating matrix  $CSI_i$  from dynamic Activities detected based on Model 4

**Output:** Generate static activities from dynamic activities, and build model 5 that adapts new location

- 1: **while** (true) **do**:
- 2:     Use the changepoint to detect changes at Equation (10) of signal  $S$  at Equation (12)
- 3:     Define logic sequence classifier to label each triggered changes in dynamic activities
- 4:     Augment the  $N_s$  samples in matrix with dimensions of  $(55 \times 200)$
- 5:     Store file as labeled activity as, sit or stand
- 6:     Train [model 5] based on labeled activities using generated data to categorize static activities into classes.
- 7:     Estimate variance changes of CSI Angle using Equation 4 and labels activities according to the generated dataset
- 8:     Input the CSI dataset  $H_i$  into the LSTM model for processing
- 9:     Obtain the predicted class label from the LSTM model output for each static activity (such as, sit/stand)
- 10:    Classify the activity as running; otherwise, classify the activity as walking.
- 11: **end**

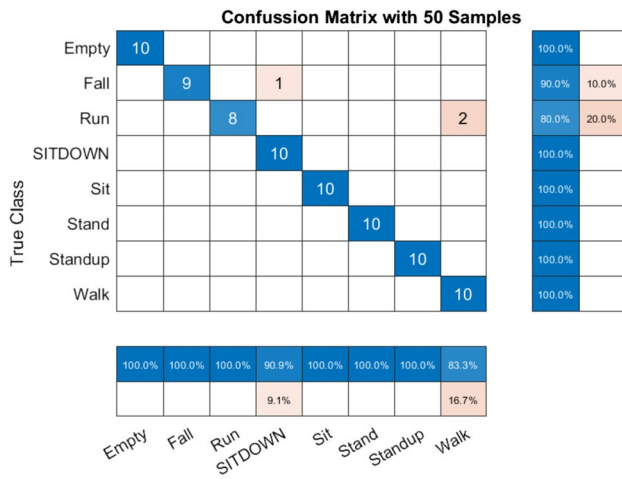


Fig. 18 Performance of model for lower samples of dataset

Table 2 Evaluation of the performance in five locations

Location	Accuracy	Location dimensions
Lab hall	97%	9×6 m
Library-hall	92%	30×18 m
Home	91%	4×3 m
Lab	92.2	20×16 m
Classroom	91	14×8 m

## 5 Results and Evaluation

The experimental setup and configuration settings are presented, followed by an assessment of the overall feasibility and effectiveness of the system. Thirdly, the individual modules comprising the system are scrutinized. Subsequently, the system’s robustness is assessed by analyzing the impact of diverse data samples. Finally, the evaluation of real-time classification and limitations are presented.

### 5.1 Experimental Setup and Layout

For experimentation, we utilized a Network Interface Card (NIC) Broadcom BCM43455c0, which supports the IEEE802.11n/ac standard with Multi-User Multiple Input Multiple Output (MU-MIMO) and is suitable for frequency bands of 20 MHz, 40 MHz, and 80 MHz. This work focused on testing the 20 MHz and 80 MHz frequency bands. For this experiment, we used the Raspberry Pi operating system

Table 3 Model performance evaluation

Model	Accuracy (%)
LSTM	97.2
Bi_LSTM	98.4
GRU	93.4
CNN	92.1
CNN-LSTM	94



Fig. 19 Testing model at untrained environments

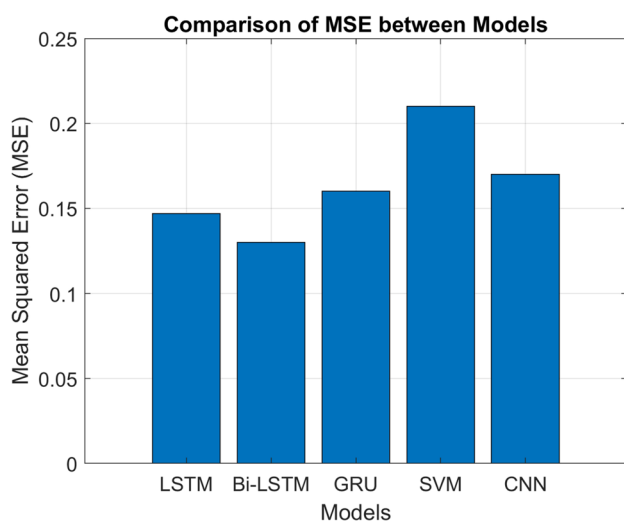


version Raspbian Buster Kernel v5.109 on a Pi 4B. The RPi captures CSI data and is set to monitor mode at 20 MHz with 5 GHz using a transmitter (Tx) (AC1350 TP-LINK router) and a receiver (Rx) (RPi 4B), both of which use omnidirectional antennas. In this section, we present the overall accuracy using samples collected at 200 frames per second, utilizing 56 subcarriers for 20 MHz bandwidth and 232 subcarriers for 80 MHz bandwidth. The model training procedure was carried out in a layout of the environment shown in Fig. 16 and then tested at the different environmental structures and varying distances between the transmitter (Tx) and receiver (Rx).

## 5.2 Overall Performance

### 5.2.1 Performance at Trained Location

We evaluated the performance of the proposed model through several stages engaged with capturing sufficient files for each activity in the designated location and subsequently assessing their accuracy. We started by capturing data for dynamic activities, and the same number of files were generated for static activities. Two frequencies, 2.4GHz and 5GHz, were utilized to further evaluate the model's efficacy, with bandwidths of 20 and 80 MHz respectively. The resulting evaluation was then compared between the two frequencies. As depicted in Fig. 17, the model trained on the higher resolution dataset, based on 80 MHz, exhibited superior performance. Nonetheless, it is essential to note that using a higher-resolution dataset requires more processing time due to its larger size.



**Fig. 20** Comparison of mean squared error (MSE) between different models

In addition to evaluating the model's performance with a more significant number of samples, it is also essential to evaluate the performance when the model is trained with few samples. One analysis evaluated the performance of a model trained with only 50 samples per activity. The results showed that the model achieved an accuracy of 96.25%, indicating that the model was able to perform well even with a limited number of training samples. One explanation for the stable accuracy with fewer samples is that the LSTM algorithm learns high-level features that are robust to variations in the data. This allows the model to generalize well to new data even when limited training data is available (Fig. 18).

### 5.2.2 Performance Across Different Environments

A set of experiments were conducted to evaluate the cross-environment performance of the proposed model. The first experiment involved training the model using a dataset collected in the Lab room and testing it in four different environments. During this experiment, the trained model was utilized to run in real time and generate new static datasets for static activities. The results of these experiments are presented in Table 2.

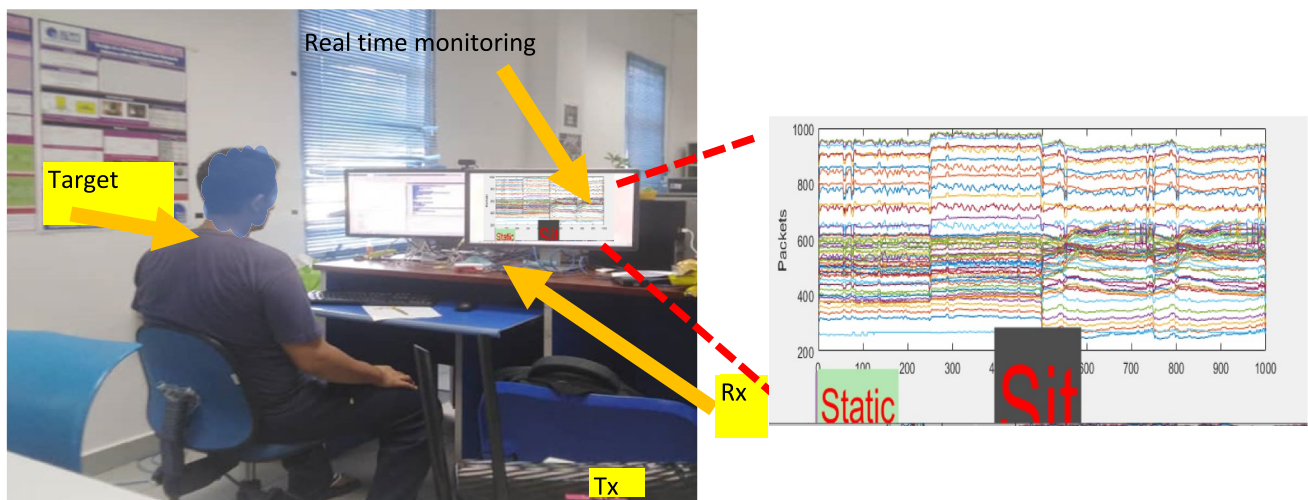
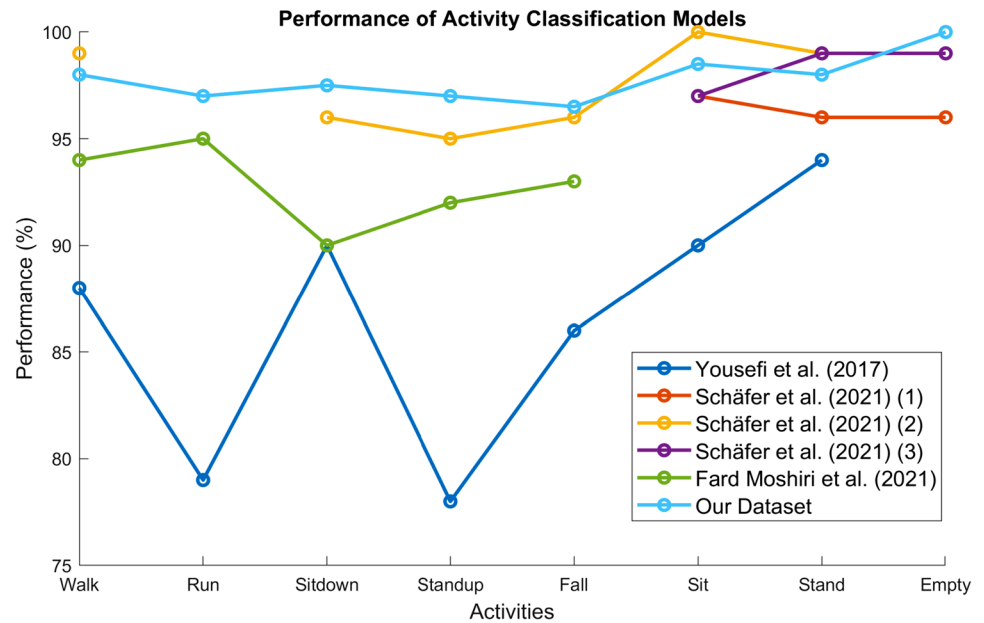
The model trained using the dataset collected in the classroom achieved an impressive accuracy of 97% in the trained environment and accuracy of approximately 92% in the other untrained four locations. These findings indicate that the proposed model could classify labeled activities based on the concatenating sequence and maintain stable performance even when the environment changes as shown in Fig. 19. This suggests that the proposed model has the potential to be deployed in various environments and still maintain high levels of accuracy in activity recognition and fall detection.

## 5.3 Module Study

### 5.3.1 Trained Model Evaluation

In order to further evaluate the performance of the proposed model, various types of neural networks were employed, including RNNs) with LSTM, bi-LSTM, GRU, CNN, and CNN-LSTM. The performance of each model was compared to determine the optimal architecture for the task at hand. It was found that the LSTM model provided stable accuracy for the trained data, mainly when dealing with sequential activities and real-time classification. This is due to the ability of LSTM to handle long-term dependencies in the data. On the other hand, the bi-LSTM and GRU models did not significantly improve accuracy compared to the LSTM model. The CNN and CNN-LSTM models, while showing high accuracy in certain situations, were found to be less effective in handling sequential data (Table 3).

**Fig. 21** Performance analysis using an available public dataset for selective activities [10, 27, 37]



**Fig. 22** Real-time monitoring of activities

Generally, a BiLSTM is slower than a unidirectional LSTM due to its processing of the input sequence in both forward and backward directions. This bidirectional nature requires more computational resources and time, making BiLSTM more computationally expensive than unidirectional LSTM. We have utilized smaller layers in the model design to optimize computational efficiency, as described in the experimental settings. The results indicate that the proposed LSTM model is well-suited for activity recognition and fall detection, particularly in real-time applications involving sequential data analysis. Despite the potential computational trade-off, the performance benefits of LSTM in capturing long-term dependencies and accurately

classifying activities justify its usage in these scenarios. The model's ability to handle sequential data with high precision and real-time responsiveness highlights its efficacy in practical applications.

By leveraging the strengths of LSTM, we achieve reliable activity recognition and fall detection, paving the way for enhanced performance measures and an improved understanding of human behavior. Additionally, The mean square error (MSE) was calculated to compare the performance of LSTM, BiLSTM, GRU, SVM, and CNN models shown in Fig. 20. The results showed that the RNN-based methods (LSTM, BiLSTM, and GRU) outperformed SVM and CNN regarding MSE, indicating their superiority in

capturing the sequential nature of activities. Furthermore, the RNN methods, with their inherent recurrent connections, have a distinct advantage in extracting relevant features from the complex and dynamic CSI data commonly encountered in activity recognition tasks. Therefore, RNN-based approaches are preferred for sequential activities and feature extraction tasks, particularly when utilizing CSI data.

### 5.3.2 Public Database Evaluation

One of the significant challenges in WiFi CSI-based activity recognition is the need for publicly available standard datasets. While there are a few datasets available, such as StanWiFi [37], SignFi [21], HuAc [2, 3, 12], and CSI data based on images [23], these datasets are limited in scope and complexity. As previously stated, researchers employ various architectures in their studies, such as the one used by [2], focusing on gesture recognition and utilizing 32 subcarriers and a 5300 NIC. However, when comparing this approach to that of [27], which focuses on activity behavior recognition and uses 64 subcarriers extracted using RPi Nexmon, it becomes clear that these architectures are not compatible in terms of hardware and activities. This incompatibility may result in poor performance when applied to other activity recognition datasets.

The proposed model utilizes public datasets to track and analyze human activity over time and exhibits high adaptability to various environments with minimal adjustments. The evaluation of the trained model with a comparable public dataset reveals its excellent performance in accurately recognizing and classifying similar activities. Specifically, Fig. 21 provides a comprehensive summary of the evaluation metrics, including precision which demonstrates the robustness and generalizability of our approach.

Upon analyzing the works of [6, 36], the proposed models primarily concentrate on location-dependent sensing with a specific emphasis on dynamic activities. In contrast, the present model demonstrates the capacity to accurately classify static activities, irrespective of the user's location, by mapping their activities in varying settings and generating datasets encompassing diverse orientations and positions. It enhances the robustness of the proposed models and broadens their potential applications beyond dynamic activity recognition.

### 5.4 Real-Time Classification Evaluation

This model presents a real-time model that accurately classifies pre-determined activities at the instance level. The model employs a Raspberry Pi 4B to continuously monitor CSI data and generate PCAP format files. A deep learning model uses MATLAB for activity classification to process

the generated files. The data undergoes a series of steps to ensure optimal performance, including receiving, decoding, preprocessing, and classifying. The process considers the available computation hardware capabilities and implements strategies to reduce data injection while achieving smooth propagation. The model's performance was assessed in multiple environments, and the results indicate high accuracy in activity classification. Figure 22 illustrates the real-time monitoring process, depicting the classification results alongside labeled predicted activities, providing clear insights into the model's performance.

## 6 Limitations and Future Works

The current implementation of WiFi CSI-based for HAR with location independence has limitations that need to be addressed. Specifically, the proposed model was designed only for certain activities, and different activities require separate analyses for each individual activity. Additionally, the model does not consider gesture recognition, which requires a different strategy for analysis and recognition. Although the presented model performs well in the initial stage of reading activities and enables the model to match activities to one another. The model is designed for a single user, and multiple sensing still needs to be improved. The beamforming could enhance the multiple sensing capabilities of the model. Model transformer models could also be another approach to recognizing activities for more than one person. Therefore, there is a need for further research and development to improve the proposed model's capabilities and overcome the challenges in location-independent sensing of human activities.

## 7 Conclusions

This work introduced a novel HAR system that facilitates location-independent sensing through an adaptive learning algorithm. The proposed system requires minimal effort to train in new locations and uses data acquired through location-free sensing inspired by sequential activities learning. The system employs the LSTM feature representation and a metric learning-based human activity mapping and recognition system to identify activities. Furthermore, the model extracts discriminative features for conditioning based on common characteristics of different locations. We evaluated the system's performance on a comprehensive dataset and found that it achieved an average accuracy of 97% for trained indoor locations and 92% at untrained locations. Additionally, the system adapts to the user's activity speed with a small amount of self-augmented data. This feature allows the system to generalize to new locations and users with

minimal effort, which is crucial for practical deployment. Therefore, based on our results, we firmly conclude that the approach is feasible and robust for location-independent sensing.

**Acknowledgements** The authors acknowledge the technical and financial support by the Ministry of Higher Education, Malaysia, under the research grant no. FRGS/1/2020/ICT02/UTEM/02/1 and Universiti Teknikal Malaysia Melaka (UTeM)

**Author contributions** Conceptualisation, Fahd Abuhoureyah and Yan Chiew Wong; methodology, software and validation, formal analysis and investigation. Fahd Abuhoureyah and Ahmad Sadhiqin; resources and data curation, Fahd Abuhoureyah and Yan Chiew Wong and Mohammed Al-Andoli; writing original draft preparation, writing—review and editing, Fahd Abuhoureyah and Yan Chiew Wong; visualisation, Yan Chiew Wong.; super- vision, Yan Chiew Wong. and Ahmad Sadhiqin.; project administration. All authors have read and agreed to the published version of the manuscript.

**Funding** This work was supported by Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education, Malaysia.

**Availability of data and materials** The data used to support the findings of this study can be obtained by emailing the corresponding author.

## Declarations

**Conflict of interest** All authors have declare that they have no conflict of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmed HFT, Ahmad H, Narasingamurthi K, Harkat H, Phang SK. DF-WiSLR: device-free wi-fi-based sign language recognition. *Pervasive Mob Comput.* 2020;69: 101289. <https://doi.org/10.1016/j.pmcj.2020.101289>.
- Alazrai R, Awad A, Alsaify B, Hababeh M, Daoud MI. A dataset for Wi-Fi-based human-to-human interaction recognition. *Data Brief.* 2020;31: 105668. <https://doi.org/10.1016/j.dib.2020.105668>.
- Alsaify BA, Almazari MM, Alazrai R, Daoud MI. A dataset for Wi-Fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments. *Data Brief.* 2020;33: 106534. <https://doi.org/10.1016/j.dib.2020.106534>.
- Ashleibta AM, Taha A, Khan MA, Taylor W, Tahir A, Zoha A, Abbasi QH, Imran MA. 5G-enabled contactless multi-user presence and activity detection for independent assisted living. *Sci Rep.* 2021;11(1):1–15. <https://doi.org/10.1038/s41598-021-96689-7>.
- Ding X, Jiang T, Zhong Y, Huang Y, Li Z. Wi-Fi-based location-independent human activity recognition via meta learning. *Sensors.* 2021;21(8):2654. <https://doi.org/10.3390/s21082654>.
- Ding X, Jiang T, Zhong Y, Huang Y, Li Z, Zhang Y, Yin Y, Wang Y, Ai J, Wu D, Ding X, Jiang T, Zhong Y, Wu S, Yang J, Gu Z, He T, Wang Z, Xu Y, et al. Wi-Fi-based location-independent human activity. *Comput Commun.* 2022;71(November 2022):87–95. <https://doi.org/10.1016/j.comcom.2022.10.027>.
- Ding X, Jiang T, Zhong Y, Wu S, Yang J. Wi-Fi-based location-independent human activity recognition with attention mechanism enhanced method. *Electronics.* 2022;11:642.
- Du H, Jin T, He Y, Song Y, Dai Y. Segmented convolutional gated recurrent neural networks for human activity recognition in ultra-wideband radar. *Neurocomputing.* 2020;396:451–64. <https://doi.org/10.1016/j.neucom.2018.11.109>.
- Elujide I, Feng C, Shiran A, Li J, Liu Y. Location independent gesture recognition using channel state information. In: *Proceedings—IEEE consumer communications and networking conference, CCNC*, pp. 841–846. 2022. <https://doi.org/10.1109/CCNC49033.2022.9700590>.
- Fard Moshiri P, Shahbazian R, Nabati M, Ghorashi SA. A CSI-based human activity recognition using deep learning. *Sensors (Basel, Switzerland).* 2021;21(21):1–19. <https://doi.org/10.3390/s21217225>.
- Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst.* 2017;28(10):2222–32. <https://doi.org/10.1109/TNNLS.2016.2582924>.
- Guo L, Wang L, Liu J, Zhou W, Lu B. HuAc: human activity recognition using crowdsourced WiFi signals and skeleton data. *Wirel Commun Mob Comput.* 2018. <https://doi.org/10.1155/2018/6163475>.
- Hao Z, Duan Y, Dang X, Liu Y, Zhang D. Wi-sl: contactless fine-grained gesture recognition uses channel state information. *Sensors (Switzerland).* 2020;20(14):1–26. <https://doi.org/10.3390/s20144025>.
- He Y, Chen Y, Hu Y, Zeng B. WiFi vision: sensing, recognition, and detection with commodity MIMO-OFDM WiFi. *IEEE Internet Things J.* 2020;7(9):8296–317. <https://doi.org/10.1109/JIOT.2020.2989426>.
- Hussain Z, Sheng QZ, Zhang WE. A review and categorization of techniques on device-free human activity recognition. *J Netw Comput Appl.* 2020;167(June): 102738. <https://doi.org/10.1016/j.jnca.2020.102738>.
- Jiang H, Cai C, Ma X, Yang Y, Liu J. Smart home based on WiFi sensing: a survey. *IEEE Access.* 2018;6:13317–25. <https://doi.org/10.1109/ACCESS.2018.2812887>.
- Li C, Liu M, Cao Z. WiHF: gesture and user recognition with WiFi. *IEEE Trans Mob Comput.* 2022;21(2):757–68. <https://doi.org/10.1109/TMC.2020.3009561>.
- Liu J, Liu H, Chen Y, Wang Y, Wang C. Wireless sensing for human activity: a survey. *IEEE Commun Surv Tutor.* 2020;22(3):1629–45. <https://doi.org/10.1109/COMST.2019.2934489>.
- Liyanage M, Chang C, Srirama S, Loke S. Indoor people density sensing using Wi-Fi and channel state information. *Adv Model Ana A.* 2018;61(1):37–47. [https://doi.org/10.18280/ama\\_b.610107](https://doi.org/10.18280/ama_b.610107).
- Lu Y, Lv S, Wang X. Towards location independent gesture recognition with commodity WIFI devices. *Electronics (Switzerland).* 2019;8(10):1–20. <https://doi.org/10.3390/electronics8101069>.
- Ma Y. Improving Wifi sensing and networking with channel state information information. *Dissertations, theses, and masters*



- projects (2020). <https://doi.org/10.21220/s2-dwgg-3j27>. <https://scholarworks.wm.edu/cgi/viewcontent.cgi?article=6877&context=etd>.
22. Ma Y, Arshad S, Muniraju S, Torkildson E, Rantala E, Doppler K, Zhou G. Location- and person-independent activity recognition with WiFi, deep neural networks, and reinforcement learning. *ACM Trans Internet Things*. 2021;2(1):1–25. <https://doi.org/10.1145/3424739>.
  23. Memmesheimer R, Theisen N, Paulus D. Gimme signals: discriminative signal encoding for multimodal activity recognition. In: *IEEE international conference on intelligent robots and systems*, March. pp. 10394–10401. 2020. <https://doi.org/10.1109/IROS45743.2020.9341699>.
  24. Muhammad G, Alshehri F, Karray F, El Saddik A, Alsulaiman M, Falk TH. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. *Inf Fusion*. 2021;76(November 2020):355–75. <https://doi.org/10.1016/j.inffus.2021.06.007>.
  25. Niu K, Zhang F, Wang X, Lv Q, Luo H, Zhang D. Understanding WiFi signal frequency features for position-independent gesture sensing. *IEEE Trans Mob Comput*. 2022;21(11):4156–71. <https://doi.org/10.1109/TMC.2021.3063135>.
  26. Oguntala GA, Hu YF, Alabdullah AAS, Abd-Alhameed RA, Ali M, Luong DK. Passive RFID module with LSTM recurrent neural network activity classification algorithm for ambient-assisted living. *IEEE Internet Things J*. 2021;8(13):10953–62. <https://doi.org/10.1109/JIOT.2021.3051247>.
  27. Schäfer J, Barriwal BR, Kokkharova M, Adil H, Liebehenschel J. Human activity recognition using csi information with nexmon. *Appl Sci (Switzerland)*. 2021;11(19):8860. <https://doi.org/10.3390/app11198860>.
  28. Sharma L, Chao C, Wu SL, Li MC. High accuracy wifi-based human activity classification system with time-frequency diagram cnn method for different places. *Sensors*. 2021;21(11):3797. <https://doi.org/10.3390/s21113797>.
  29. Shi Z, Zhang JA, Xu R, Cheng Q, Pearce A. Towards Environment-independent human activity recognition using deep learning and enhanced CSI. In: *2020 IEEE global communications conference, GLOBECOM 2020—proceedings*. 2020. <https://doi.org/10.1109/GLOBECOM42002.2020.9322627>
  30. Suthar B, Gadhia B. Human activity recognition using deep learning: a survey. *Lect Notes Data Eng Commun Technol*. 2021;52:217–23. [https://doi.org/10.1007/978-981-15-4474-3\\_25](https://doi.org/10.1007/978-981-15-4474-3_25).
  31. Wang D, Yang J, Cui W, Xie L, Sun S. Multimodal CSI-based human activity recognition using GANs. *IEEE Internet Things J*. 2021;8(24):17345–55. <https://doi.org/10.1109/JIOT.2021.3080401>.
  32. Wang F, Liu J, Gong W. WiCAR: WiFi-based in-car activity recognition with multi-adversarial domain adaptation. In *Proceedings of the International Symposium on Quality of Service, IWQoS 2019*. 2019. <https://doi.org/10.1145/3326285.3329054>
  33. Wang Y, Zheng Y. Modeling RFID signal reflection for contact-free activity recognition. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2018;2(4):1–22. <https://doi.org/10.1145/3287071>.
  34. Yang J, Liu Y, Liu Z, Wu Y, Li T, Yang Y. A framework for human activity recognition based on WiFi CSI signal enhancement. *Int J Antennas Propag*. 2021. <https://doi.org/10.1155/2021/6654752>.
  35. Yang J, Zou H, Jiang H, Xie L. Fine-grained adaptive location-independent activity recognition using commodity WiFi. In: *IEEE wireless communications and networking conference, WCNC, 2018-April*, pp. 1–6. 2018. <https://doi.org/10.1109/WCNC.2018.8377133>.
  36. Zhang Y, Liu Q, Wang Y, Yu G. CSI-based location-independent human activity recognition using feature fusion. *IEEE Trans Instrum Meas*. 2022;71:1–17. <https://doi.org/10.1109/TIM.2022.3216419>. (Id. 3216419).
  37. Yousefi S, Narui H, Dayal S, Ermon S, Valaei S. A survey of human activity recognition using WiFi CSI. pp. 1–8. <http://arxiv.org/abs/1708.07129>.
  38. Zeng Y, Wu D, Xiong J, Zhang D. Boosting WiFi sensing performance via CSI ratio. *IEEE Pervasive Comput*. 2021;20(1):62–70. <https://doi.org/10.1109/MPRV.2020.3041024>.
  39. Zhai S, Tang Z, Wang D, Li Q, Li Z, Chen X, Fang D, Chen F, Wang Z. Enhancing received signal strength-based localization through coverage hole detection and recovery. *Sensors (Switzerland)*. 2018;18(7):16–22. <https://doi.org/10.3390/s18072075>.
  40. Zhang Y, Yin Y, Wang Y, Ai J, Wu D. CSI-based location-independent human activity recognition with parallel convolutional networks. *Comput Commun*. 2023;197(November 2022):87–95. <https://doi.org/10.1016/j.comcom.2022.10.027>.
  41. Zhang Y, Zheng Y, Qian K, Zhang G, Liu Y, Wu C, Yang Z. Widar3.0: zero-effort cross-domain gesture recognition with Wi-Fi. *IEEE Trans Pattern Anal Mach Intell*. 2021. <https://doi.org/10.1109/TPAMI.2021.3105387>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.