

Received 5 September 2023, accepted 20 September 2023, date of publication 27 September 2023, date of current version 2 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320069

## RESEARCH ARTICLE

# Efficient Wi-Fi-Based Human Activity Recognition Using Adaptive Antenna Elimination

MIR KANON ARA JANNAT<sup>1</sup>, MD. SHAFIQL ISLAM<sup>1</sup>, SUNG-HYUN YANG<sup>1</sup>,  
AND HUI LIU<sup>2</sup>, (Member, IEEE)

<sup>1</sup>Smart H&B Technology Laboratory, Department of Electronic Engineering, Kwangwoon University, Seoul 01890, Republic of Korea

<sup>2</sup>Cognitive Systems Laboratory, University of Bremen, 28359 Bremen, Germany

Corresponding authors: Sung-Hyun Yang (shyang@kw.ac.kr) and Hui Liu (hui.liu@uni-bremen.de)

This work was supported in part by the University of Bremen; in part by Deutschen Forschungsgemeinschaft (DFG) via Universitätsbibliothek Bremen (SuUB) Bremen; in part by the Ministry of Trade, Industry and Energy, Republic of Korea, through the AI Home Platform Development Project, under Grant 20009496; and in part by the Excellent Researcher Support Project of Kwangwoon University, in 2023.

**ABSTRACT** Recently, Wi-Fi-based human activity recognition using channel state information (CSI) signals has gained popularity due to its potential features, such as passive sensing and adequate privacy. The movement of various body parts in between Wi-Fi signals' propagation path generates changes in the signal reflections and refraction, which is evident from the CSI variations. In this paper, we analyze the relationship between human activities and properties (amplitude and phase) of Wi-Fi CSI signals on multiple receiving antennas and discover the signal properties that vary remarkably in response to human movement. The variation in the signal received among multiple antennas shows different sensitivity to human activities, directly affecting recognition performance. Therefore, to recognize human activities with better efficiency, we proposed an adaptive antenna elimination algorithm that automatically eliminates the non-sensitive antenna and keeps the sensitive antennas following different human activities. Furthermore, the correlation of the statistical features extracted from the amplitude and phase of the selected antennas' CSI signal was analyzed, and a sequential forward selection was utilized to find the best subset of features. Using such a subset, three machine learning algorithms were employed on two available online datasets to classify various human activities. The experimental results revealed that even when using easy-to-implement, non-deep machine learning, such as random forest, the recognition system based on the proposed adaptive antenna elimination algorithm achieved a superior classification accuracy of 99.84% (line of sight) on the StanWiFi dataset and 97.65% (line-of-sight) / 93.33% (non-line-of-sight) on another widely applied multi-environmental dataset at a fraction of the time cost, demonstrating the robustness of the proposed algorithm.

**INDEX TERMS** Human activity recognition, channel state information, adaptive antenna elimination, machine learning, artificial intelligence, Wi-Fi, smart home.

### LIST OF ACRONYMS

**AAE** adaptive antenna elimination  
**CNN** convolutional neural network  
**CSI** channel state information  
**FN** false negative  
**FP** false positive  
**GRU** gated recurrent unit  
**HAR** human activity recognition

**HMM** hidden Markov model  
**KNN**  $k$ -nearest neighbor  
**LOS** line-of-sight  
**LSTM** long short-term memory  
**MAD** mean absolute deviation  
**MIMO** multiple-input multiple-output  
**ML** machine learning  
**NLOS** non-line-of-sight  
**OFDM** orthogonal frequency-division multiplexing  
**PCC** Pearson correlation coefficients

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

<b>PWR</b>	power
<b>RF</b>	random forest
<b>RSSI</b>	received signal strength indicator
<b>SFS</b>	sequential forward selection
<b>SNR</b>	signal-to-noise ratio
<b>SVM</b>	support vector machine
<b>TN</b>	true negative
<b>TP</b>	true positive

## I. INTRODUCTION

Human Activity Recognition (HAR) is one of the key factors in a smart home that monitors daily activities using a set of Internet of Things gadgets in the context of a smart home. Thanks to such monitoring, a smart house can provide occupants, especially disabled and aging citizens, with individualized home care services to enhance their quality of life, independence, and health. HAR has gradually become one of the most notable research topics in diverse fields because of its numerous applications, including fall detection [37], [54], elderly monitoring [36], context awareness [21], [34], driver behavior analysis [25], sports assistance [5], gesture recognition [39], edutainment [18], and information retrieval [12], [40], among others. The sophisticated design and suitable usage of sensors are essential to obtain high-quality sensory observations that may be used to identify users' activities and behaviors [30].

Conventional HAR methods use different sensing technologies: vision-based [1], wearable-based [47], and radars [27]. Vision-based is a prominent and first sensing technology where a camera is used to obtain human movement data from the neighboring environment, which achieves good performance but is susceptible to various factors, such as measuring volume, light and environmental conditions, and user privacy. These limitations significantly impact the outcome of vision-based HAR in many applications. In wearable-based HAR, signals are obtained from wearable devices and sensors, which has been widespread and fruitful. In addition, sometimes users need to carry different devices, for example, smartphones, smartwatches, and smart bracelets, to gather cumbersome human activity data, which is inconvenient, particularly for elders and disabled people. Moreover, activity recognition could be impracticable if the person does not have the required tools with them. Radar-based HAR collects data using specialized equipment, such as universal software radio peripherals, but its coverage area is limited.

Wi-Fi-based methods, hybridizing the following features, can be the emerging way to bridge the issues associated with vision, sensor, and radar sensing technologies [51]. First, no camera/wearable and non-invasiveness enable uncomplicated data acquisition while preserving adequate privacy. Second, Wi-Fi signals have extensive transmission that can propagate through walls, furniture, and doors, ubiquitous in indoor environments as a reliable and vital source of information. Third, Wi-Fi signals are reflected by the human

body [24], [48], whose changing patterns can be utilized for detecting different activities. Last but not least, the ease and efficiency of Wi-Fi signal acquisition allow it to be potentially applicable to real-time or online HAR [18], a task that often requires complex research when utilizing vision or wearable technologies, particularly against the capacity, speed, and accuracy of data transmission [33].

Currently, available Wi-Fi-based HAR uses two metrics to represent a Wi-Fi signal: received signal strength indicator (RSSI) and channel state information (CSI). RSSI represents coarse-grained information and has been used primarily for indoor localization [20] and HAR [13]. RSSI signal analyzes changes in the received signal's strength but fails to detect the changing signals generated by a moving person. As the gap between human movement and the receiver antenna increases, the RSSI signal, which measures changes in the strength of the received signal, becomes erroneous, affecting the overall performance of the system [35]. In contrast, CSI measures fine-grained signals and is applied to a variety of tasks, including handwriting recognition [15], position estimation [49], fall detection [53], and micromovement identification [50]. CSI signals expediently capture the propagation signal information between the transmitter and receiver antenna pairs at a particular carrier frequency. As summarized in [37], generic human activities can be classified based on CSI signals using time-frequency features such as torso and limb velocities by characterizing the changing speed of the reflected path length in subcarrier amplitudes. Such velocities are similar to those obtained from Doppler radars without movement direction. The critical insight of HAR through the Wi-Fi CSI signal is that the amplitude and phase of the CSI signal generated by human movement or object differ from the regular measurement without movement, through which we can analyze the received signals, in particular their amplitude and phase information, to recognize human activities. Furthermore, existing research demonstrates that CSI signals perform better in complicated situations than RSSI (see Section II) because, in interior environments, the amplitude and phase information provided by CSI can clearly distinguish between moving and steady signal patterns. This article applies Wi-Fi CSI signals for the HAR study.

There are challenges associated with employing CSI signals to recognize human activities. The relationship between various human movements, surrounding environments, and antenna placement results in different antenna sensitivities to different activities. Antennas are easily affected by external factors due to changing environment and different human activities. Hence, in this article, an adaptive antenna elimination algorithm is proposed to eliminate nonsusceptible antennas and keep the most sensitive antenna signals related to human activities. Moreover, several statistical features are computed from each selected antenna's amplitude and phase. The correlation of the extracted features is analyzed, and to reduce the data overfitting, the wrapper feature selection technique has been used to find the most informative features.

Three machine learning (ML) models are trained with the optimal feature set on two open datasets, evaluated by a ten-fold cross-validation. The experimental results demonstrated that the proposed HAR performs better than or on par with state-of-the-art methods for recognizing human activity with respect to accuracy, precision, recall, and F1-score, with a significant enhancement of efficiency and simplicity. The main contributions of this article are:

- We proposed an adaptive antenna elimination algorithm to automatically eliminate non-sensitive antennas and keep critical antennas, which reduces unnecessary data and enables efficient HAR;
- Unlike previous studies that solely employ either amplitude or phase information, we jointly utilize the amplitude and the phase information in the CSI sequence to enhance the recognition accuracy;
- Our proposed system, erected on a set of experimentally selected superior features, outperforms the existing state-of-the-art works on two widely-applied datasets in terms of its high-level activity recognition capability with a low time cost.

## II. BACKGROUND AND RELATED WORK

Because of the availability and numerous advantages of Wi-Fi communication, Wi-Fi signal-based HAR methods have become popular recently. In this section, we review existing Wi-Fi-based HAR methods, where those existing methods are divided into RSSI and CSI.

### A. HUMAN ACTIVITY RECOGNITION BASED ON RECEIVED SIGNAL STRENGTH INDICATOR (RSSI)

RSSI signal-based HAR approaches rely on the change in the received signal strength introduced by different human movements. Sigg et al. [43] designed a HAR approach in which a mobile phone collected RSSI signals to recognize human activities. They acquired data from the three cases of a mobile phone: lying on a table in an empty room, lying on a table while a subject is moving around the room, and being held and managed by a subject. They extracted and selected features, based on which the system achieved recognition accuracy of 52% for 11 gestures and 72% for four chosen gestures.

An online HAR system by Gu et al. [14] analyzed RSSI fingerprints from various human activities. Gu et al. [13] presented a unique recognition case utilizing the RSSI signal and developed a fusion approach with  $k$ -nearest neighbor to identify human activities at an average accuracy of 92.58%. A hardware device was applied in Sigg et al. [44] to obtain RSSI signals from the surrounding environment for the recognition of four different activities, including walking, standing, crawling, and lying down, which reported an accuracy greater than 80%. Youssef et al. [57] presented a localization system using an RSSI signal, which detects different environmental changes and tracks passive entities.

So far, RSSI has been less effective in complex operations since it only provides coarse information about channel

variations and can often be influenced by multipath effects and noise [7], [51].

### B. CHANNEL STATE INFORMATION (CSI)

CSI describes the channel characteristics of a communication link, referring to the propagation of the signal affected by human movement between the transmitter and receiver to indicate distance, scattering, the effect of power, and fading, among others. CSI can be applied in cognitive scenarios such as HAR, gesture recognition, and location tracking by virtue of its distinguished sensitivity to the environment. Modern wireless technology adopts the multiple-input multiple-output (MIMO) system consisting of multiple transmitting and receiving antennas. Each transmitter-receiver antenna pair forms a communication channel to send adjacent information over the established channel using various modulation techniques, the most widespread of which is orthogonal frequency-division multiplexing (OFDM) that uses a MIMO channel's bandwidth to send information on several concurrent orthogonal subcarrier frequencies. Each of these subcarriers can be identified using CSI. In CSI metric/wireless systems, the MIMO-OFDM technique can be modeled as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{v}, \quad i = 1, 2, 3, \dots, N, \quad (1)$$

where  $\mathbf{H}_i \in \mathbb{C}^{N_R \times N_T}$  represents the CSI matrix of the  $i^{\text{th}}$  subcarrier,  $\mathbf{v}$  denotes the noise term,  $N$  denotes the number of OFDM subcarriers, and  $\mathbf{y}_i \in \mathbb{R}^{N_{R_x}}$  and  $\mathbf{x}_i \in \mathbb{R}^{N_{T_x}}$  are the  $i^{\text{th}}$  received and transmitted signals, respectively.

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

where  $h_i^{jk}$  is the CSI of the  $i^{\text{th}}$  subcarrier for the link between the  $j^{\text{th}}$  transmitted antenna and the  $k^{\text{th}}$  receiving antenna.

Using the CSI matrix  $\mathbf{H}_i$ , CSI describes the signal attenuation factor along each transmission line. This includes signal scattering, power decay with distance, multipath fading, and other details. The amplitude-frequency and phase-frequency properties of a signal are used by the channel frequency response to characterize the multipath propagation of the signal. The mathematical expression of the frequency response is given as

$$h_i^{jk} = |h_i^{jk}| e^{j \angle h_i^{jk}}, \quad (3)$$

where  $|h_i^{jk}|$  and  $\angle h_i^{jk}$  denote amplitude and phase, respectively.

### C. HUMAN ACTIVITY RECOGNITION BASED ON CSI

Compared to RSSI, which measures only power over the entire channel bandwidth, CSI provides a set of channel estimates for each subcarrier of each transmission link.

Wang et al. [52] designed a CSI-based human action recognition and monitoring method (CARM) using time-frequency features such as torso and limb velocities. CARM consists of two models: a CSI-speed model describing the connection between CSI kinematics and human motions and a CSI-activity model explaining the relationship between human activities and movement speed. CARM performed a recognition rate of 96% and can withstand environmental changes. Yousefi et al. [56] created the StanWiFi dataset from Wi-Fi-based CSI signals, on which statistical features were extracted for the modeling research of long short-term memory (LSTM), hidden Markov model (HMM), and random forest. Chen et al. [7] proposed an attention-based bidirectional LSTM (ABLSTM) model to identify passive human activities utilizing CSI signals, outperforming other benchmarks. A Wi-Motion system was proposed in [26] to recognize five activities in line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios. Moshiri et al. [11] presented a deep learning-based HAR approach on Raspberry Pi by collecting CSI signals of seven human activities. They transformed CSI signals into 2-D images using pseudocolor plots and evaluated four convolutional neural networks (CNN) or LSTM models, namely 1D-CNN, 2D-CNN, LSTM, and bidirectional LSTM, among which 2D-CNN reached 95% accuracy. Alsaify et al. [2] constructed a multi-environmental HAR system using CSI signals at an overall recognition rate of 91.27% through denoising, activity segmentation, statistical feature extraction and selection, and support vector machine training. Salehinejad and Valaee [41] introduced a LiteHAR model with a random convolution kernel for feature extraction, achieving 93% accuracy on a public dataset. Yadav et al. [55] presented an advanced version of the InceptionTime network, named CSITime, and evaluated it on three public datasets. Shalaby et al. [42] studied four deep learning models on the StanWiFi dataset, among which the CNN gaited recurrent unit (CNN-GRU) model acquired an accuracy of 99.31% with a time cost of 0.0033 seconds per sample. The accuracy of the attention model on CNN-GRU is 99.16%, but the sample-wise time consumption drops to 0.0019 seconds. Very recently, Islam et al. [22] proposed a deep learning model, called spatio-temporal convolution with nested LSTM (STC-NLSTMNet), which reaches 99.88% and 98.20% accuracies on two public datasets.

The literature listed above delivers that several researchers have worked on recognizing human activity using diverse Wi-Fi signal-based approaches in signal processing, ML, and deep learning. One issue that has not yet received enough attention is that the mapping of human activities with Wi-Fi CSI signals on multiple antennas is affected by different sensitivities due to the fact that the Wi-Fi signals received by the antenna can be reflected during the transmission. The study on sensitivity changes in response to different activities in a multi-antenna scenario, including amplitude and phase variations, should help enhance HAR, motivated by which we analyzed the antenna's sensitivity following different activities and proposed an adaptive antenna elimination

(AAE) algorithm that can eliminate non-sensitive antenna data during signal processing.

AAE aspires to reduce complexity and improve training and recognition efficiency by eliminating the least sensitive antenna. Pursuing the model simplification is also a reasonable consideration in the spirit of further reducing complexity. We found that most Wi-Fi-based HAR studies are based on deep learning, which is known to be more complex and time-consuming than regular ML in principle. Therefore, three non-deep ML models are applied in this article, aiming to

- Confirm whether AAE is effective with plain ML models rather than deep learning. The latter might naturally lead to better results;
- Provide a lightweight Wi-Fi-based HAR scheme, which not only minimizes redundancy in data utilization but also engages minimalism in the model, thus ensuring simplicity, low consumption and high efficiency, and partial interpretability.

### III. DATASETS

Data corpus is an essential and critical resource in all aspects of ML and deep learning. We employed two publicly available Wi-Fi signal-based datasets for our HAR study:

#### STANWIFI

Yousefi et al. collected a dataset in an LOS environment of an indoor area [56]. We refer to this dataset as StanWiFi, following the name given in [41] and [55]. Six participants performed each of the six activities (Fall, Run, Lie down, Walk, Sit down, and Stand up) twenty times. The experimenters used one Wi-Fi router with one antenna as a transmitter and a laptop equipped with an Intel® 5300 NIC with three antennas as a receiver to collect data. The receiver was located three meters from the transmitter, and the sampling rate was 1000 Hz. Each subject completed each session within 20 seconds, starting and ending in a stationary state.

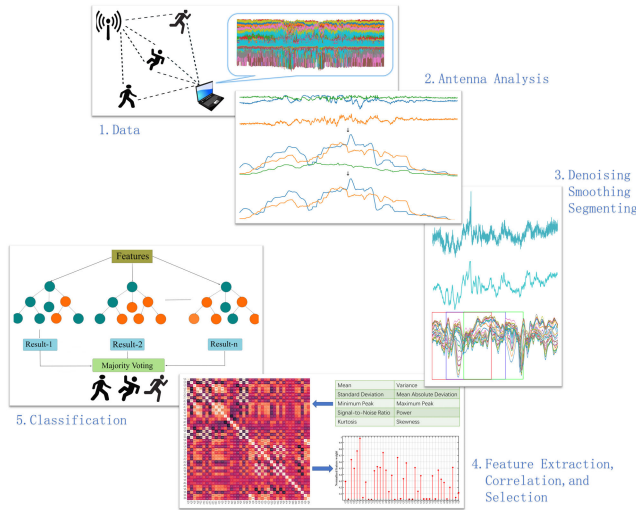
#### MULTIENVIRONMENT

The dataset [3] was collected from three indoor environments. The provider did not name this dataset; for the sake of narrative, we will call it "MultiEnvironment" in this article. One of the LOS scenarios, the office, set the distance between the receiver and the transmitter at 3.7 meters. The NLOS scenario had a barrier (an eight-centimeter wall) between the transmitter and the receiver. Thirty persons were asked to perform five sessions: falling from a standing or sitting position, walking, sitting down and standing up from a chair, and picking a pen from the ground, each repeated 20 times. Subsequently, each session was divided into several activities. For example, falling from a sitting position was divided into three activities: sitting, falling, and lying. The authors of [2] and [4], who belong to the group of dataset owners, identified 12 classes from these five sessions and reorganized them into six labels for HAR, listed in Table 1. We followed the data provider's work by applying the six-category activity



**TABLE 1. Human activities and their descriptions in the “MultiEnvironment” dataset [3].**

Activity	Description
No movement	Sitting, standing, or lying on the ground
Falling	Falling from a standing position or a chair
Sitting down / Standing up	Sitting down on a chair or standing up from a chair
Walking	Walking between the transmitter and the receiver
Turning	Turning at the transmitter or at the receiver
Picking up a pen from the ground	—



**FIGURE 1. Overall workflow of the proposed Wi-Fi signal-based HAR system.**

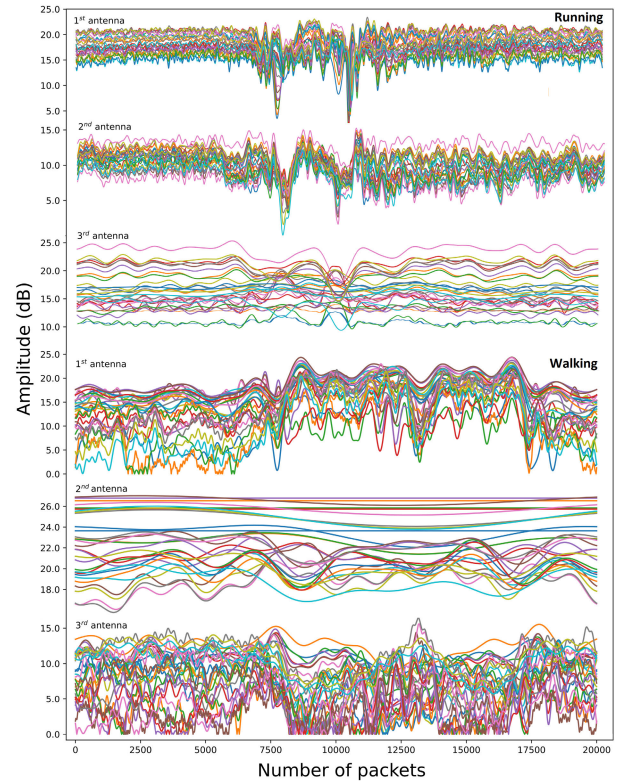
labeling in our study on both the LOS (office) and the NLOS scenarios. The total number of samples is 3,000 (30 subjects × 5 experiments × 20 trials). Two computers were used to collect data, one as a one-antenna transmitter and the other as a three-antenna receiver.

**IV. METHODS**

Figure 1 portrays the overall workflow of the proposed HAR system in this article, which consists of five primary aspects: data preparation, antenna analysis, signal processing, feature research, and modeling.

**A. ANTENNA ANALYSIS FOR ADAPTIVE ELIMINATION**

A MIMO system consists of multiple transmitter and receiver antennas. Both the StanWiFi and “MultiEnvironment” datasets have one transmitting and three receiving antennas. The Wi-Fi signals from the transmitting antenna can be reflected by the human body during the propagation route to the receiving antenna. In the context of the variety of human movements and the surroundings, antennas are sensitive to external information like human movement direction and the antenna’s vertical dimension. As a result, receiving antennas have different susceptibilities to different activities, for which previous works concentrated on subcarrier selection and fusion techniques [46]. However, subcarrier selection on non-sensitive antennas does not show any significance, which indicates that various antennas exhibit varying sensitivity to the same human activity.



**FIGURE 2. Example of three antennas’ signal visualization for the activities running (top) and walking (bottom). Each antenna has 30 subcarriers indicated in different colors.**

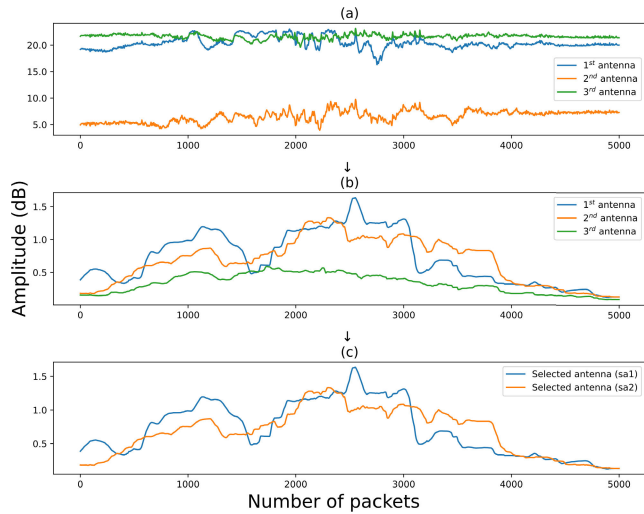
Figure 2 (Running) diagnoses that for the activity running in such a case, the first and second antennas are more sensitive, while the third does not show significant perceptibility. Similarly, the first and third antennas exhibit more sensitivity for walking, as Figure 2 (Walking) evinces. It can be witnessed that the amplitude signal of a sensitive antenna changes substantially, while that of a non-sensitive antenna keeps relatively stable. Hence, eliminating non-sensitive antennas should be a potential way to enhance HAR.

The signal that the non-sensitive antenna receives is severely corrupted by the surrounding noise and only vaguely depicts human activities. Possible reasons are a combination of factors such as environments, antenna positioning, and human body movement. Our goal is to distinguish the sensitive and non-sensitive antennas. Therefore, we proposed an adaptive antenna elimination algorithm that adaptively eliminates antennas based on their sensitivity to different human activities:

- 1) Let  $CSI_{a,s,p}(i)$  denote the  $i^{th}$  sample’s CSI value in packet  $p$  of antenna  $a$ ’s subcarrier  $s$ . Antenna  $a$ ’s mean value sequence of all its subcarriers is computed as

$$\mu_a(p) = \overline{\sum_{s \in a} \sum_{i \in p} CSI_{a,s,p}(i)}, \tag{4}$$

where in our applied datasets,  $a = 1, 2, \text{ or } 3$  and  $s = 1, 2, \dots, 30$ , respectively.



**FIGURE 3.** Example of an adaptive antenna elimination (AAE) procedure on a CSI data stream: (a) average values of each antenna's 30 subcarriers; (b) standard deviation values based on the signals in (a) using sliding window; (c) the elimination of the non-sensitive antenna based on values in (b).

- 2) To distinguish the antenna's sensitivity to human activities, the sliding window-based *standard deviation sequence* of antenna  $a$ 's  $\mu_a(p)$  is calculated with a window size  $l$ . The  $n^{\text{th}}$  standard deviation value in the sequence is

$$\sigma_a(n) = \sqrt{\frac{(\sum_{i=n}^{n+l-1} (\mu_a(i) - \sum_i \mu_a(i))^2)}{l-1}}, \quad (5)$$

where in this work,  $l$  and the sliding step equal 3 and 1, respectively. Let  $\sigma_a = [\sigma_a(1), \sigma_a(2), \dots]$  denote the *standard deviation sequence*.

- 3) Antenna  $a$ 's *sensitivity* is the maximum minus the minimum of its  $[\sigma_a]$ :

$$\text{sensitivity}_a = \max \sigma_a - \min \sigma_a, \quad (6)$$

based on which the antenna with the lowest sensitivity should be eliminated.

The procedure elucidated above provides a reasonable basis for selecting between sensitive and non-sensitive antennas, as displayed in Figure 3 (b), a visualization perspective preliminarily validating the possibility of eliminating non-sensitive antennas: The proposed AAE algorithm eliminates one antenna based on the lowest values and keeps the other two antennas, as resulted in Figure 3 (c). The pseudocode of AAE is given in algorithm 1.

In the current version of AAE, two of the three antennas are retained and one is eliminated. Questions may arise. What if all three antennas are sensitive, one is sensitive, or all are non-sensitive? More sophisticated selections can be made, but at the current stage they are not necessary because

- If a selection algorithm screens out three non-sensitive antennas, no data is left to use. In this case, it makes sense to eliminate the least sensitive one and keep two to provide some helpful information;

### Algorithm 1 Pseudocode of Adaptive Antenna Elimination (AAE)

- 1: **Input:** CSI data sequence  $D$  containing multiple antennas with multiple subcarriers each
- 2: **Output:** Remaining antennas after eliminating the least sensitive antennas
- 3: **function** AAE ( $D$ ):
- 4:   window size  $l \leftarrow 3$
- 5:    $step \leftarrow 1$
- 6:   **for**  $a$  in all antennas **do**
- 7:     mean value sequence  $\mu_a \leftarrow$  mean values of antenna  $a$ 's data sequence  $D(a)$  (see Equation 4)
- 8:     window no.  $n \leftarrow 0$
- 9:     standard deviation sequence  $\sigma_a \leftarrow [ ]$
- 10:     **while**  $n + l \leq \text{Length}(D(a))$  **do**
- 11:       sliding window  $w(n) \leftarrow \mu_a[n : n + l]$
- 12:       calculate the standard deviation  $\sigma(w(n))$  (see Equation 5)
- 13:       append  $\sigma(w(n))$  to  $\sigma_a$
- 14:        $n \leftarrow n + step$
- 15:     **end while**
- 16:     Sensitivity $_a = \max \sigma_a - \min \sigma_a$  (see Equation 6)
- 17:   **end for**
- 18:   eliminate antenna  $\arg \min_a \text{Sensitivity}_a$ ; return the rest
- 19: **end function**

- If a selection algorithm determines that all three antennas are sensitive, it is evidenced from numerous signal visualizations that the three antennas have a high mutual correlation. Eliminating one should have minimal effect on the results. In our work, it even improves the recognition performance (see Table 7);
- If a selection algorithm judges only one antenna to be sensitive, keeping one of the other two insensitive ones may provide additional information, evidenced in Table 7, while the negative impact of the least sensitive one is reduced;
- Most importantly, excellent recognition rates were achieved in the experiments (see Sections V-D and V-E) using the current AAE algorithm that eliminates one non-sensitive antenna. Therefore, we do not propose a more sophisticated elimination scheme at this stage.

This work aims to validate the novel idea of antenna elimination for HAR by confirming whether the classifier performs well when eliminating the most insensitive antenna. After successful validation, future tasks include designing reasonable threshold metrics to eliminate or retain more antennas, especially for datasets that contain a large number of antennas.

### B. DENOISING, SMOOTHING, AND SEGMENTATION

Based on the AAE approach, the antenna that is calculated as less sensitive to human activities is eliminated. Since Wi-Fi-based CSI signals are intervened by signal attenuation,

**TABLE 2.** Feature applied in this study.  $x_1, x_2, \dots, x_N$  is a sequence of  $N$  samples for feature extraction;  $a_1$  and  $a_2$  indicate the two selected antennas.

Feature	Mathematical Expression	Notation	
		Amplitude ( $A$ )	Phase ( $P$ )
Mean	$\mu = \frac{\sum_{i=1}^N x_i}{N}$	$F(a_1, \mu, A)$	$F(a_1, \mu, P)$
		$F(a_2, \mu, A)$	$F(a_2, \mu, P)$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}}$	$F(a_1, \sigma, A)$	$F(a_1, \sigma, P)$
		$F(a_2, \sigma, A)$	$F(a_2, \sigma, P)$
Variance	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}$	$F(a_1, \sigma^2, A)$	$F(a_1, \sigma^2, P)$
		$F(a_2, \sigma^2, A)$	$F(a_2, \sigma^2, P)$
Maximum	$M = \max(x_1, x_2, \dots, x_N)$	$F(a_1, M, A)$	$F(a_1, M, P)$
		$F(a_2, M, A)$	$F(a_2, M, P)$
Minimum	$m = \min(x_1, x_2, \dots, x_N)$	$F(a_1, m, A)$	$F(a_1, m, P)$
		$F(a_2, m, A)$	$F(a_2, m, P)$
Kurtosis	$\kappa = \frac{\sum_{i=1}^N (x_i - \mu)^4}{(N-1)\sigma^4}$	$F(a_1, \kappa, A)$	$F(a_1, \kappa, P)$
		$F(a_2, \kappa, A)$	$F(a_2, \kappa, P)$
Skewness	$sk = \frac{\sum_{i=1}^N (x_i - \mu)^3}{(N-1)\sigma^3}$	$F(a_1, sk, A)$	$F(a_1, sk, P)$
		$F(a_2, sk, A)$	$F(a_2, sk, P)$
Signal-to-noise ratio	$SNR = \frac{\mu}{\sigma}$	$F(a_1, SNR, A)$	$F(a_1, SNR, P)$
		$F(a_2, SNR, A)$	$F(a_2, SNR, P)$
Power	$PWR = \frac{1}{N} \sum_{i=1}^N x_i^2$	$F(a_1, PWR, A)$	$F(a_1, PWR, P)$
		$F(a_2, PWR, A)$	$F(a_2, PWR, P)$
Mean absolute deviation	$MAD = \frac{1}{N} \sum_{i=1}^N  x_i - \mu $	$F(a_1, MAD, A)$	$F(a_1, MAD, P)$
		$F(a_2, MAD, A)$	$F(a_2, MAD, P)$

multipath propagation, and other environmental factors, it is necessary to remove noise, outliers, and other irrelevant information from raw CSI data to improve the credibility of activity-related signal patterns. The most common Butterworth bandpass filter denoises the raw CSI signal, expressed by the following equations:

$$|H_{\text{low-pass}}(j\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\omega_0})^{2n}}}. \tag{7}$$

$$|H_{\text{high-pass}}(j\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\omega_0})^{-2n}}}. \tag{8}$$

After denoising, some burrs on the human activity signal envelopes may still exist because of surrounding and equipment impacts, for which the Gaussian smoothing function further diminishes the interference of irrelevant information:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \tag{9}$$

The signal becomes cleaner after denoising and smoothing, which guarantees reliable feature extraction to enhance training and classification. Furthermore, segmentation separates signals into smaller pieces, commonly referred to as windows or frames, which helps resolve constraints during data preprocessing with the aim of training and recognition [23], [32]. The first difficulty is that the data from different participants may have different lengths of recorded trials. Another problem is that processing a long time series consumes resources significantly. Our experiments applied a window size of 512 samples with a stride of 64 samples (12.5% overlap ratio) to segment the processed CSI signal.

**C. FEATURE EXTRACTION, CORRELATION, AND SELECTION**

Feature extraction plays a crucial role in ML. Publicly available code libraries, such as TSFEL [6], assist in the extraction of time-series features in the temporal, statistical,

and spectral domains. For each window of 512 samples (see Section IV-B), 10 handcrafted feature types (see Table 2) were extracted on the amplitude and phase of each of the two selected antennas, bringing the dimension of the feature vectors to 40.

The extracted features related to every human activity are not always correlated with each other. Correlation is a metric which measures the similarity between two features. If two features are linearly dependent, the correlation coefficient value lies between -1 and +1, while no correlation is indicated with a zero. Pearson’s correlation coefficients (PCC) [8] analyze the extracted features.

Figure 4 details the feature correlation matrix for both the amplitude and phase of the CSI signal. The matrix’s entries display the PCC between each pair of features. Some features are highly correlated, suggesting that they should be essential for recognition tasks.

Some earlier work used feature space reduction to enable the training and recognition of HAR with decreasing dimensionality [17]. Feature selection is another way to comprehensively study the decline in the number of features for HAR, which cuts down on computational expenditures and potentially improves recognition performance. Additionally, it tends to decrease overfitting and facilitate data visualization. In this study, sequential forward selection (SFS) [9], a wrapper approach based on a greedy top-down search algorithm, generated the optimal feature set. The algorithm begins with an empty subset, adds the feature that resulted in the most significant improvement in each iteration, and halts when the performance of the validation set cannot be further improved.

The importance of features was computed in terms of weights using SFS, as illustrated in Figure 5. Seventeen features do not significantly affect classification (weight < 0.2), and the remaining features are considered advantageous. Empirically and experimentally, features with a



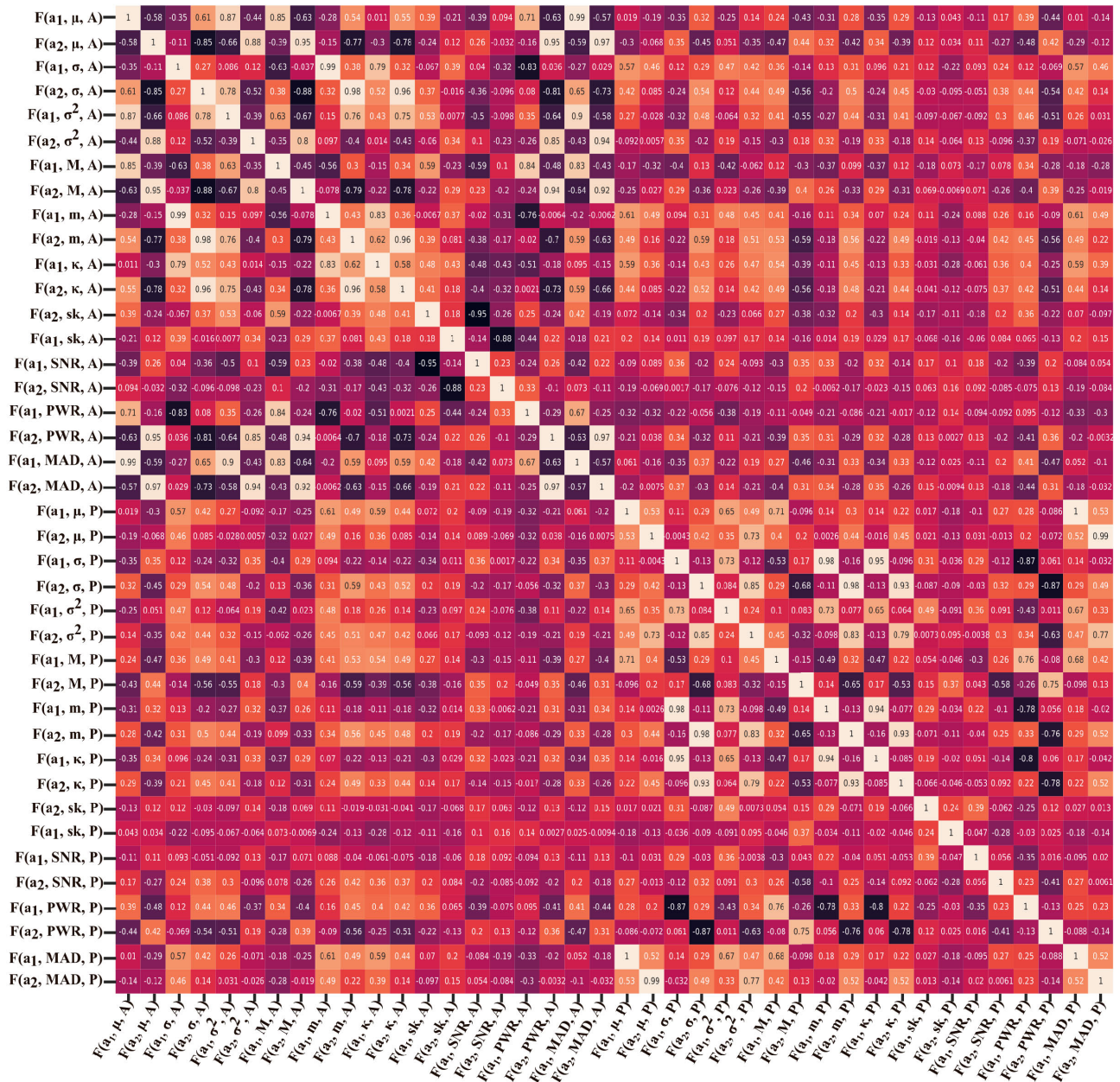


FIGURE 4. Matrix of the Pearson correlation coefficients (PCC) between each pair of features.

weight greater than 0.4 perform well. We selected fourteen features from the two antennas, specifically, ten features from amplitude and four from phase, reducing the use of handcrafted features by 65% according to the feature selection experiments.

**D. MACHINE LEARNING MODELS**

The random forest (RF) is based on ensemble learning and the collection of decision trees constructed using portions of all data [28], forecasting hypotheses from each tree to finally predict according to the majority vote. Generally,

RF minimizes the overfitting problem, and as the number of learners increases, the generalization errors decrease.

The support vector machine (SVM) is a set of supervised learning methods characterized by kernel functions [45], which works based on a hyperplane dividing an  $n$ -dimensional space into different data classes. The maximal margin determines the ideal hyperplane, although the margin itself refers to the distance between two support vectors. The data points closest to the hyperplane are the support vectors and are referred to as critical points.

The  $K$ -nearest neighbor (KNN) stores all available data and classifies a new sample according to the  $K$  nearest



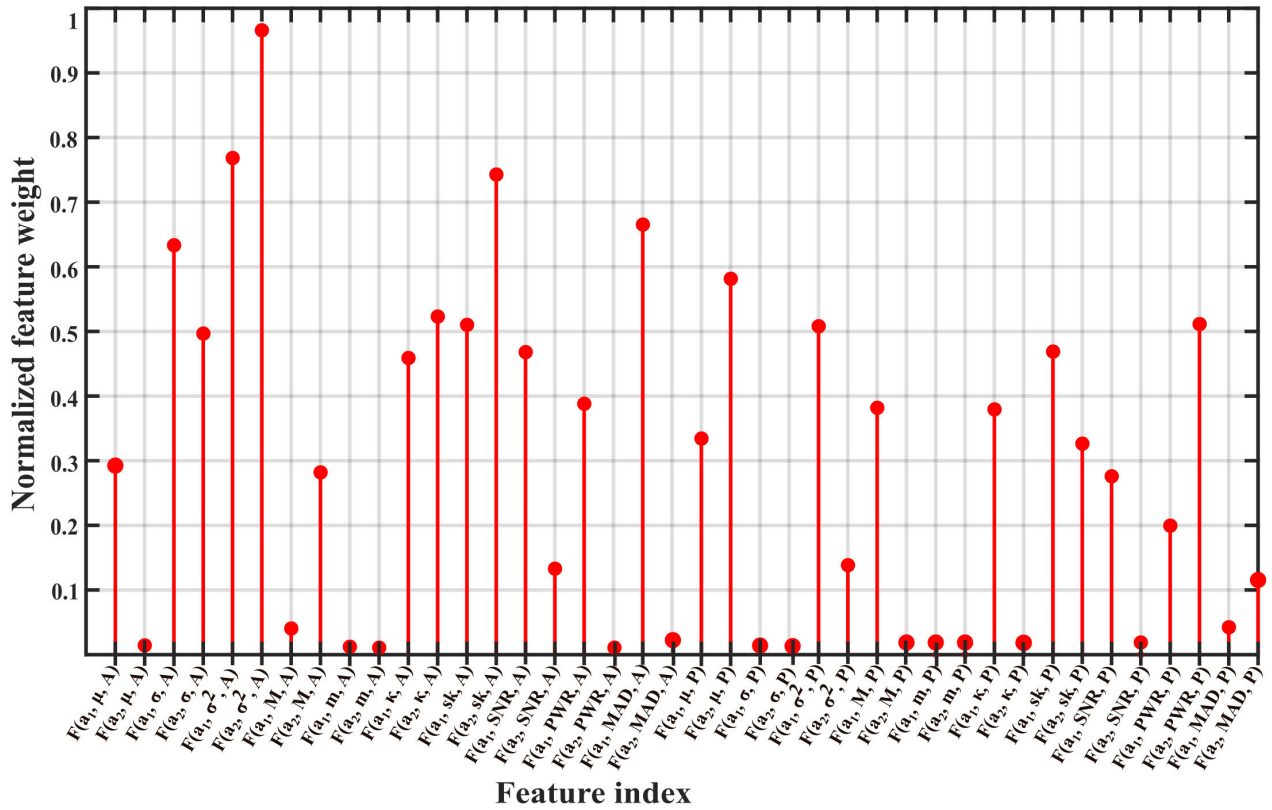


FIGURE 5. Importance of the extracted features for HAR.

samples by calculating different types of distances [10], such as the Euclidean. KNN handles both linear and non-linear scenarios, capable of unknown morphological data.

This work used all the default settings of the above-mentioned three ML models in the *Python* package *scikit-learn* [38].

E. EVALUATION

We used a ten-fold cross-validation procedure, which randomly splits the whole dataset into ten non-overlapping parts. Each experimental session took nine folds for training, and the rest one fold for recognition and evaluation. We have evaluated the overall recognition performance of the proposed system using well-known performance evaluation metrics, such as accuracy, precision, recall, and F1-score, based on the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

TABLE 3. AAE-based HAR results on the StanWiFi dataset. To keep more valid digits, the precision, recall, and F1-score values are expressed using percentages. The percent sign is omitted from all statistics.

Classifier	Metrics	Fold										Average
		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	10 <sup>th</sup>	
RF	Accuracy	99.76	99.83	99.93	99.86	99.93	99.83	100.00	99.93	99.86	99.86	99.84±0.07
	Precision	98.98	99.88	99.89	99.99	100.00	99.89	99.97	99.83	99.88	99.99	99.82±0.30
	Recall	99.89	99.99	100.00	99.87	99.88	99.83	98.99	99.99	99.85	99.97	99.83±0.30
	F1-score	99.43	99.93	99.94	99.91	100.00	100.00	99.04	99.95	100.00	99.98	99.81±0.32
SVM	Accuracy	97.27	97.24	97.20	96.89	97.26	96.99	97.28	97.33	97.33	97.30	97.20±0.15
	Precision	96.98	96.99	97.10	97.19	97.20	97.28	97.29	97.16	97.27	97.23	97.16±0.11
	Recall	96.89	97.20	97.13	97.18	97.19	97.30	97.20	97.18	97.10	97.33	97.17±0.12
	F1-score	96.93	97.09	97.11	97.17	97.19	97.28	97.23	97.16	97.18	97.27	97.16±0.10
KNN	Accuracy	95.99	95.78	95.89	95.68	95.98	95.83	95.89	95.83	96.87	97.14	96.08±0.49
	Precision	95.81	95.88	95.85	95.25	95.64	95.79	95.90	95.83	96.87	97.22	96.00±0.59
	Recall	95.91	95.89	95.85	95.78	95.78	95.69	95.97	95.73	96.51	97.36	96.04±0.52
	F1-score	96.06	95.81	95.88	95.75	95.91	95.82	95.26	95.61	96.75	97.62	96.05±0.67

V. RESULTS AND DISCUSSION

The results of the AAE-based HAR experiments on the two datasets were summarized and supplemented with a comparison of the state-of-the-art results.

A. EXPERIMENTS ON THE STANWIFI DATASET

Table 3 reports the experimental results of our proposed HAR system on the StanWiFi dataset, where the RF classifier achieves the highest performance with an average of more than 99.80% accuracy/F1-score/precision/recall. Among the ten-fold sessions, the first fold shows the lowest accuracy (99.76%), and the 7<sup>th</sup> obtains all correct (100%). The SVM and KNN classifiers reach average accuracies of 97.20%±0.15% and 96.08%±0.49%, respectively.

Figure 6 shows the confusion matrix of the RF classifier on the StanWiFi dataset. RF accurately classifies the activities

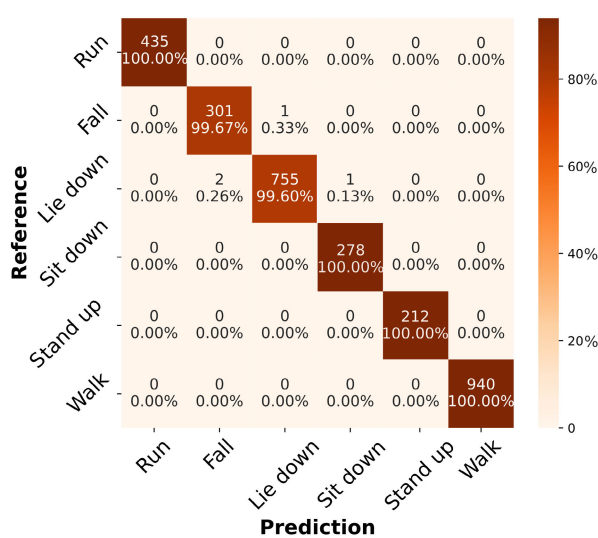


FIGURE 6. RF classifier's confusion matrix of the ten-fold AAE-based HAR results on the StanWiFi dataset.

TABLE 4. AAE-based HAR results on the "MultiEnvironment" dataset's LOS (office) scenario. To keep more valid digits, the precision, recall, and F1-score values are expressed using percentages. The percent sign is omitted from all statistics.

Classifier	Metrics	Fold										Average	
		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	10 <sup>th</sup>		
RF	Accuracy	96.99	96.94	97.85	97.72	97.86	97.83	97.75	97.55	98.18	97.84	97.84	97.65 ± 0.39
	Precision	96.97	96.90	97.83	97.71	97.85	97.79	97.71	97.55	98.15	97.83	97.63	97.63 ± 0.39
	Recall	96.98	96.93	97.84	97.70	97.84	97.80	97.73	97.52	98.05	97.82	97.62	97.62 ± 0.38
	F1-score	96.98	96.92	97.84	97.71	97.85	97.80	97.72	97.54	98.10	97.84	97.63	97.63 ± 0.38
SVM	Accuracy	93.88	94.54	94.87	94.99	95.10	95.26	95.40	95.54	95.68	95.78	95.10 ± 0.57	
	Precision	94.44	94.76	93.88	95.21	95.26	95.30	95.45	95.20	95.10	95.37	95.00 ± 0.49	
	Recall	94.99	95.01	95.10	95.11	95.23	95.34	95.35	93.78	95.10	95.36	95.08 ± 0.45	
	F1-score	94.71	94.88	94.48	95.15	95.24	95.31	95.40	94.48	95.10	95.33	95.01 ± 0.35	
KNN	Accuracy	94.99	94.98	95.06	95.98	95.93	95.97	95.94	95.99	95.99	96.89	93.88 ± 0.55	
	Precision	95.00	94.99	95.76	95.98	95.99	96.10	95.99	95.97	95.88	95.99	93.76 ± 0.41	
	Recall	94.99	95.43	95.88	95.89	95.93	95.97	95.98	96.20	95.10	95.98	93.73 ± 0.41	
	F1-score	94.99	95.20	95.81	95.93	95.96	96.03	95.98	96.08	95.48	95.90	93.74 ± 0.38	

TABLE 5. AAE-based HAR results on the "MultiEnvironment" dataset's NLOS scenario. To keep more valid digits, the precision, recall, and F1-score values are expressed using percentages. The percent sign is omitted from all statistics.

Classifier	Metrics	Fold										Average
		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	10 <sup>th</sup>	
RF	Accuracy	92.26	92.85	93.75	93.88	92.85	93.89	93.83	92.83	93.88	93.86	93.33 ± 0.62
	Precision	92.85	93.30	92.78	93.63	93.10	93.30	92.08	93.05	93.55	93.51	93.12 ± 0.49
	Recall	91.21	92.55	93.68	93.51	93.36	92.90	93.72	93.15	93.17	93.39	93.07 ± 0.74
	F1-score	91.60	92.50	93.52	93.64	93.47	93.67	92.61	93.28	93.63	93.52	93.14 ± 0.69
SVM	Accuracy	88.98	89.99	88.95	88.96	89.97	87.98	85.10	86.89	87.97	89.99	88.47 ± 1.56
	Precision	88.68	89.82	88.86	88.96	89.88	87.78	83.89	86.88	87.76	89.76	88.42 ± 1.34
	Recall	85.56	88.78	87.72	89.69	89.68	88.89	85.09	86.80	88.06	88.67	88.14 ± 1.38
	F1-score	88.85	87.82	88.70	88.60	89.85	88.58	86.05	85.89	88.46	88.88	88.16 ± 1.26
KNN	Accuracy	80.88	80.99	81.85	80.69	81.99	82.58	79.50	81.15	78.95	79.99	80.85 ± 1.14
	Precision	80.58	80.28	81.84	80.65	81.88	82.52	81.18	78.90	79.55	79.95	80.73 ± 1.13
	Recall	80.55	80.35	81.25	80.25	81.85	82.88	81.26	78.88	79.69	79.88	80.68 ± 1.16
	F1-score	80.67	80.58	81.35	80.50	81.90	82.91	81.00	78.95	79.80	79.98	80.76 ± 1.12

Run, Sit down, Stand up, and Walk without failure, whereas the recognition rate is over 99% for Fall and Lie down.

**B. EXPERIMENTS ON THE "MULTIENVIRONMENT" DATASET**

Table 4 and Table 5 record the experimental results of the RF classifier in the LOS and NLOS scenarios, respectively, while Figures 7 and 8 convey the confusion matrices of the RF classifier for both scenarios. The LOS HAR tasks witness an over 97.61% overall accuracy/precision/recall/F1-score, while SVM and KNN perform 95.10%±0.57% and 93.88%±0.55% recognition rates, respectively.

It is observable from Figures 7 and 8 that Sitting down/Standing up is the key factor pulling down the

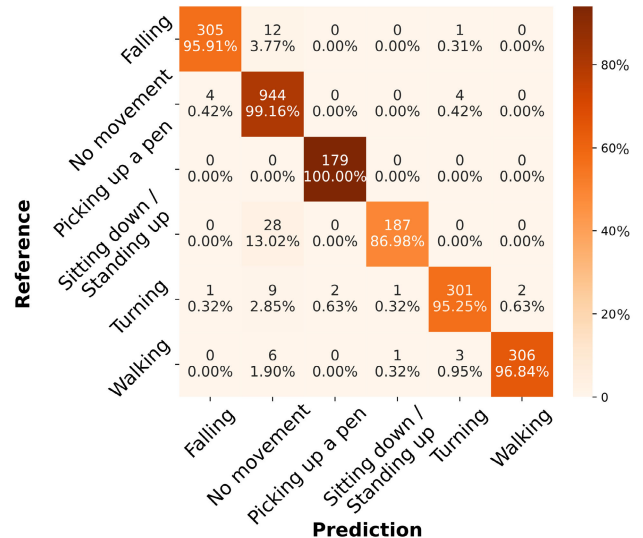


FIGURE 7. RF classifier's confusion matrix of the ten-fold AAE-based HAR results on the "MultiEnvironment" dataset's LOS (office) scenario.

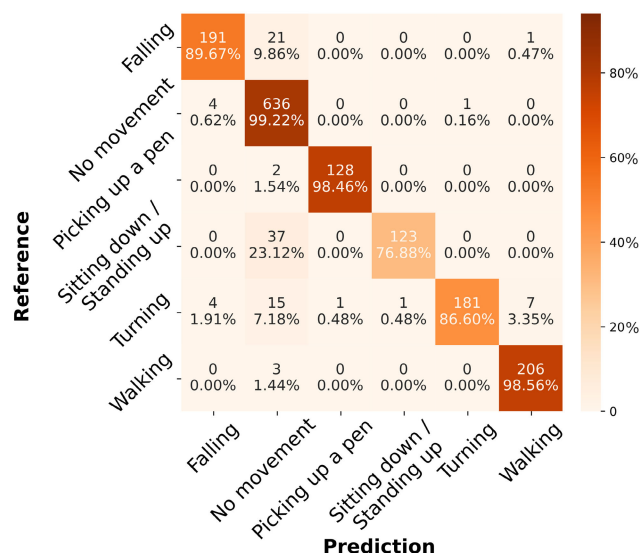


FIGURE 8. RF classifier's confusion matrix of the ten-fold AAE-based HAR results on the "MultiEnvironment" dataset's NLOS scenario.

global recognition rate, followed by Falling and Turning. Recognition errors for these three activities were mainly concentrated on No movement, for which a possible reason could be that there is a stationary state ("No movement") of the body prior to performing Sitting down, Standing up, Falling, or Turning. The application of sequential modeling like HMM may enhance universal recognition accuracy [29], [31], which should be one of the potential follow-up studies. Furthermore, it should not be overlooked that the class Sitting down/Standing up combines two activities in opposite directions, which follows the data providers' labeling (see Section III) but undoubtedly increases recognition difficulty. A strong contrast is that the StanWiFi dataset does not

**TABLE 6.** Summary of our proposed method's results with other state-of-the-art published results.

Dataset	Source	Model	Metrics (%)				Time Cost (seconds)		
			Accuracy	Precision	Recall	F1-score	Total training	Total testing	Per sample
StanWiFi	(Yousefi et al. [56], 2017)	LSTM (2017)	90	-	-	-	-	-	-
	(Chen et al. [7], 2018)	ABLSTM (2018)	97.30	-	-	-	13,007.20	6.86	-
	(Yadav et al. [55], 2022)	CSITime (2022)	98.00	99.16	98.87	99.01	-	-	-
	(Salehinejad and Valaee et al. [41], 2022)	LiteHAR (2022)	93.00	-	-	-	157.80	5.46	0.013
	(Shalaby et al. [42], 2022)	CNN-GRU (2022)	99.31	99.50	99.43	-	-	-	0.0033
	(Islam et al. [22], 2022)	STC-NLSTMNet (2022)	<b>99.88</b>	99.72	99.73	99.72	679	80	0.0028
	<b>Proposed in this article</b>	<b>AAE+RF</b>	99.84	<b>99.82</b>	<b>99.83</b>	<b>99.81</b>	<b>45.12</b>	<b>0.29</b>	<b>0.000086</b>
"MultiEnvironment" LOS (office)	(Alsaify et al. [4], 2021)	SVM (2021)	94.03	-	-	-	-	-	-
	(Alsaify et al. [2], 2022)	SVM (2022)	91.27	-	-	-	-	-	-
	(Islam et al. [22], 2022)	STC-NLSTMNet (2022)	<b>98.20</b>	<b>98.10</b>	<b>98.08</b>	<b>98.09</b>	710	87	0.0030
	<b>Proposed in this article</b>	<b>AAE+RF</b>	97.65	96.42	96.41	94.40	<b>49.20</b>	<b>0.35</b>	<b>0.000092</b>
"MultiEnvironment" NLOS	(Islam et al. [22], 2022)	STC-NLSTMNet (2022)	<b>94.68</b>	<b>94.57</b>	<b>94.55</b>	<b>94.56</b>	710	87	0.0030
	<b>Proposed in this article</b>	<b>AAE+RF</b>	93.33	93.12	93.07	93.14	<b>49.20</b>	<b>0.35</b>	<b>0.000092</b>

**TABLE 7.** Experimental results of applying 1–3 receiving antennas on the RF classifier.

Dataset	Number of Antennas	Metrics (%)				Time Cost (seconds)		
		Accuracy	Precision	Recall	F1-score	Total training	Total testing	Per sample
StanWiFi	3 (all antennas)	97.92	97.90	97.86	97.88	65.02	0.46	0.000086
	2 (AAE)	<b>99.84</b>	<b>99.82</b>	<b>99.83</b>	<b>99.81</b>	45.12	0.29	
	1 (the most sensitive)	94.06	94.18	94.06	94.05	<b>20.50</b>	<b>0.14</b>	
"MultiEnvironment" LOS (office)	3 (all antennas)	95.10	94.90	94.85	94.87	74.01	0.52	0.000092
	2 (AAE)	<b>97.65</b>	<b>97.63</b>	<b>97.62</b>	<b>97.63</b>	49.02	0.35	
	1 (the most sensitive)	91.72	91.65	91.51	91.58	<b>28.00</b>	<b>0.19</b>	
"MultiEnvironment" NLOS	3 (all antennas)	90.40	90.32	90.31	90.32	74.01	0.52	0.000092
	2 (AAE)	<b>93.33</b>	<b>93.12</b>	<b>93.07</b>	<b>93.14</b>	49.20	0.35	
	1 (the most sensitive)	88.95	88.17	88.12	88.15	<b>28.00</b>	<b>0.19</b>	

merge these two activities, enabling both to be entirely correctly recognized (see Figure 6). Direction-related high-level features may be beneficial for such a problem [16], [19].

### C. PERFORMANCE COMPARISON AND DISCUSSION

Applying different models on the same dataset is a compelling criterion to judge. We compare the performance of our proposed system with state-of-the-art peer workpieces on the same two datasets, resulting in the statistics in Table 6.

The first issue to be noted overall is that [22], a recent publication, used more sophisticated deep learning, whose ten-fold cross-validated recognition accuracies are higher than this article by 0.04%, 0.55%, and 1.33% on StanWiFi, "MultiEnvironment"-LOS, and "MultiEnvironment"-NLOS, respectively. However, such results do not affect the success of validating the proposed AAE algorithm. With a slight loss of accuracy, AAE-based non-deep learning brings not only less complexity and more interpretability but also a considerable reduction of training and recognition time. As described in Section II-C, non-deep learning models are applied in this work to validate AAE's usefulness reasonably; hence, the almost on-par accuracies and significantly lower time cost compared to deep learning confirm AAE's effectiveness and contribution to efficient, low-consuming, lightweight Wi-Fi-based HAR.

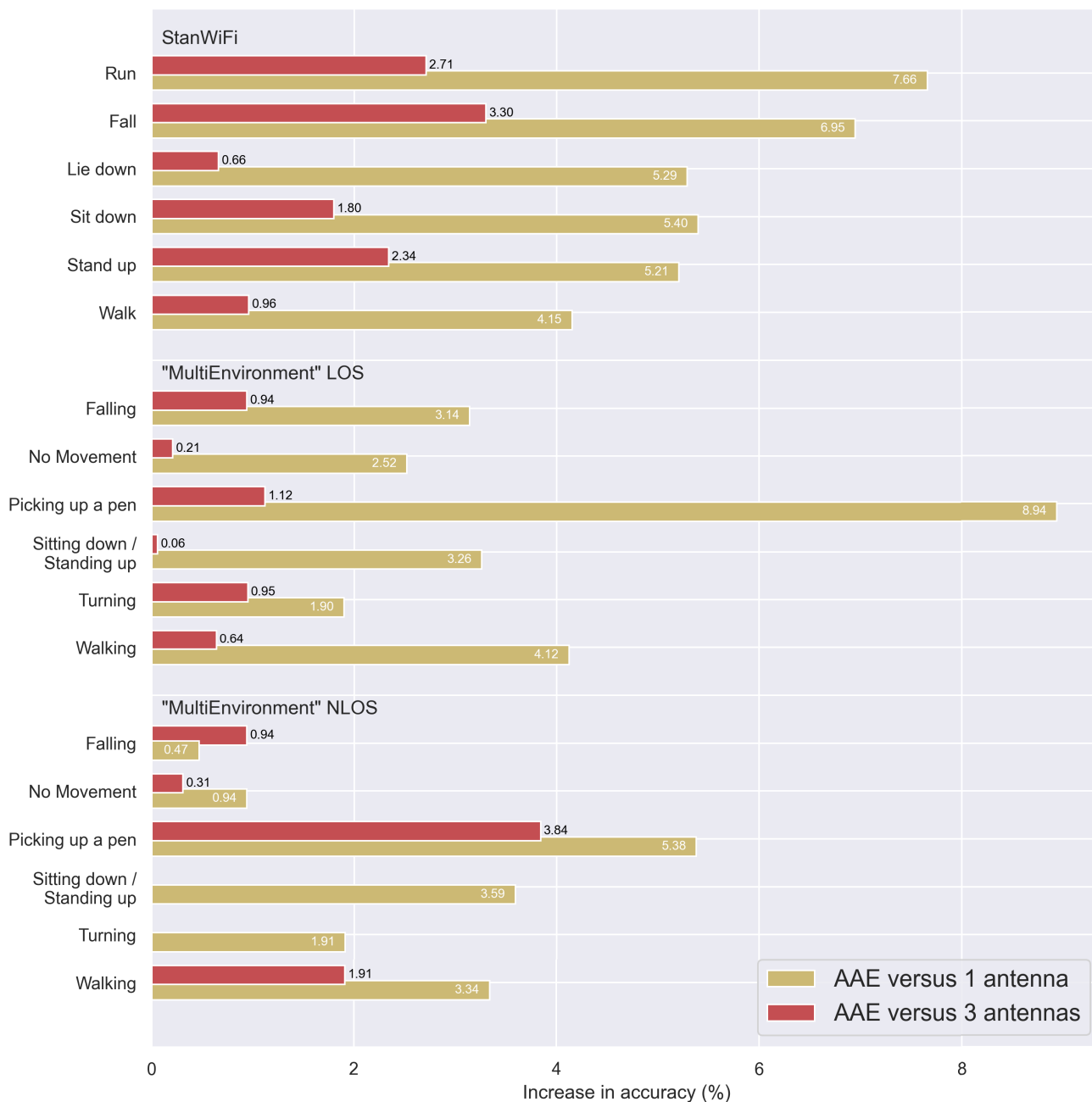
Yousefi et al. [56], the provider of the StanWiFi dataset, applied RF, HMM, and LSTM based on feature extraction to validate the dataset, among which LSTM achieved the optimal accuracy of ~90% according to [41] and [55]. Since

then, there have been several HAR studies on the StanWiFi dataset. In [7], the authors proposed the ABLSTM model that helps focus on essential features, which reached 97.30% accuracy on a random-selected ten-fold cross-validation. The advanced InceptionTime model, put forward in [55] and named CSITime, attained an accuracy of 98.00%. A lightweight HAR model (LiteHAR) [41] had a 93.00% accuracy rate on ten-fold cross-validation. In comparison, our work presented a promising recognition result of 99.84% better than or on par with most state-of-the-art works using the AAE-based RF classifier on the StanWiFi dataset, with significantly lower time consumption.

On the "MultiEnvironment" dataset, in addition to [22], there are two published referential results on the LOS scenario. In [4], features were extracted from the time and frequency domains, and the optimal set of features was selected. With SVM, an accuracy of 94.00% was reported on a 100-fold purely person-dependent experiment (ten-fold cross-validation for each of the ten subjects). In [2], after outlier removal and signal smoothing, an SVM classifier based on feature extraction and selection achieved 91.27% accuracy for LOS on a ten-fold leave-one-out cross-validation. Our work reached 97.65% accuracy on a random-selected ten-fold cross-validation in the LOS scenario.

To our knowledge, the up-to-date publicly available HAR work on the "MultiEnvironment" dataset's NLOS scenario is the recent publication [22]. The AAE-based recognition results in this work can serve as the first benchmark for non-deep learning, with comparable accuracy





**FIGURE 9.** Increase in recognition accuracy of AAE for each activity compared with using all three antennas (red) and using only the most sensitive one (yellow) on all datasets.

and significantly lower time consumption to the deep learning model.

Time complexity is a challenging factor for any model or classifier, should one desire to implement it to solve real-world problems. We also compare training and testing time costs. Our work took 45.12 seconds for training and 0.29 seconds for testing on the StanWiFi dataset, while [41] took 157.80/5.46 seconds. ABLSTM [7] spent even more. Again, the 49.20/0.35 seconds for training/testing

can be regarded as a benchmark of time cost for “MultiEnvironment”. Admittedly, the time statistics are relevant to hardware, but the efficiency of our method is evident.

The following factors contributed to the outstanding recognition performance of our work:

- The proposed AAE algorithm eliminated the antennas that are non-sensitive to human activities, significantly reducing irrelevant information.

- Statistical features were extracted from both the amplitude and the phase of the selective antennas.
- The correlation between the extracted features was analyzed using PCC. The feature selection using the SFS technique further endowed ML tasks with the most informative input.
- Following non-deep ML approaches on several publicly available datasets, AAE-based HAR achieved comparable performance to deep learning methods, with significantly reduced time complexity.

#### D. IMPACT OF THE NUMBER OF ANTENNAS RETAINED

All the datasets applied in this paper applied three receiving antennas. Therefore, the current process of eliminating one antenna, which we used to verify whether eliminating the antenna impacts recognition, implies the preservation of two antennas. To further confirm the optimality of the derived results, experiments employing all three antennas as well as retaining only the most sensitive antenna on the RF classifier, which performs best among the three applied ML models, were also performed on all datasets, as summarized in Table 7.

The advantages of AAE are self-evident in Table 7. With the elimination of one antenna, which saves one-third of the data usage, the recognition rates and F1-scores, in turn, improve by about 2–3 percentage points. When one more antenna is eliminated, that is, only the most sensitive antenna is kept, the performance drops by 5–6 percentage points. The time complexity is understandably close to proportional to the number of retained antennas, while the time cost per sample is constant. Since the time expenditure of our algorithm is already exponentially better than the state-of-the-art models (see Table 7), AAE is apparently superior to keeping only the most sensitive antenna: Doubling the time at the millisecond level brings about a significant increase in recognition rate.

The current work verifies that under reasonable algorithms, such as the proposed AAE, eliminating antenna can even bring gains to the applied datasets instead of impairing the recognition performance. When there are more antennas involved, how many antennas should be retained/eliminated to achieve the optimal accuracy-efficiency balance is a topic depending on the dataset and purpose. Future work includes designing a generalizable threshold scheme to automatically or semi-automatically determine the number of eliminated/retained antennas.

#### E. IMPACTS OF ANTENNA ELIMINATION ON EACH ACTIVITY

To explore the impact of antenna elimination on the recognition rate of each activity, Figure 9 summarizes the accuracy statistics of individual activities from the nine experiments involved in Table 7. The red bars indicate the incremental accuracies brought about by applying AAE compared to using all three antennas, while the yellow bars represent AAE's recognition rate increment compared to using only the most sensitive antenna.

The noteworthy points found in Figure 9 are listed below.

- Eliminating one least sensitive antenna, i.e., the proposed AAE, had no negative impact on any activity.
- AAE achieved a 100% recognition rate for several activities (see Figures 6–8); the other two settings never reached 100% on any individual activity.
- On the “MultiEnvironment” dataset's NLOS scenario, AAE did not gain an increase in accuracy (nor did it get worse) for Sitting down/Standing up and Turning. Besides, AAE brings only tiny gains on Sitting down/Standing up in the LOS scenario. As analyzed in Section V-B, these two activities are inherently challenging to recognize compared to the others. However, when this label is split into two separate classes, that is, in StanWiFi, AAE can perform significantly better.
- AAE earns good gains for Walking and Falling that are involved in all datasets.
- The most positively affected activity by AAE is Picking up a pen in the “MultiEnvironment” dataset's NLOS scenario. For such an activity of localized details, the elimination of an insensitive antenna greatly helped recognition. Eliminating one more antenna on top of AAE, the most considerable decrease in recognition rate was as well for Picking up a pen (in the LOS scenario), which suggests that one antenna is far from enough to characterize this activity of localized details.
- When only the most sensitive antenna was retained, the results were much worse than when all antennas or AAE were applied, with one exception: the Falling in the “MultiEnvironment” dataset's LOS scenario favors only one antenna over all antennas. Of course, AAE is still better than both.

## VI. CONCLUSION

This article presents a Wi-Fi CSI signal-based HAR system with superior HAR performance for both LOS and NLOS scenarios. An adaptive antenna elimination (AAE) algorithm has been proposed to keep the most sensitive antennas related to human activities, and various features were extracted from the amplitude and phase of the selected sensitive antennas. Feature correlation analysis and feature selection are applied to obtain the best subset of features. Three non-deep ML classifiers were applied and compared rather than deep learning models in order to confirm AAE's effectiveness generally. The proposed HAR system with RF classifiers has achieved a classification accuracy of 99.84% on the StanWiFi dataset and 97.65% (LOS) / 93.33% (NLOS) on the “MultiEnvironment” dataset, outperforming or on par with the recognition performance of state-of-the-art studies, with a significantly lower time cost.

Currently, our research involves two datasets with different types of daily activities as an AAE's proof of concept. A further investigation of our proposed system on other available datasets with more types of human activities will be a future direction. Additionally, recognizing

multi-user activities based on Wi-Fi signals is realistic but more challenging.

## REFERENCES

- [1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 1–43, 2011.
- [2] B. A. Alsaify, M. M. Almazari, R. Alazrai, S. Alouneh, and M. I. Daoud, "A CSI-based multi-environment human activity recognition framework," *Appl. Sci.*, vol. 12, no. 2, p. 930, Jan. 2022.
- [3] B. A. Alsaify, M. M. Almazari, R. Alazrai, and M. I. Daoud, "A dataset for Wi-Fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments," *Data Brief*, vol. 33, Dec. 2020, Art. no. 106534.
- [4] B. A. Alsaify, M. M. Almazari, R. Alazrai, and M. I. Daoud, "Exploiting Wi-Fi signals for human activity recognition," in *Proc. 12th Int. Conf. Inf. Commun. Syst. (ICICS)*, May 2021, pp. 245–250.
- [5] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *Proc. 23th Int. Conf. Archit. Comput. Syst.*, Feb. 2010, pp. 1–10.
- [6] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "TSFEL: Time series feature extraction library," *SoftwareX*, vol. 11, Jan./Jun. 2020, Art. no. 100456.
- [7] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wi-Fi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.
- [8] I. Cohen, Y. Huang, J. Chen, J. Benesty, J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, pp. 1–4.
- [9] S. F. Cotter, B. D. Rao, K. Kreutz-Delgado, and J. Adler, "Forward sequential algorithms for best basis selection," *IEE Proc. Vis., Image, Signal Process.*, vol. 146, no. 5, pp. 235–244, 1999.
- [10] P. Cunningham and S. J. Delany, "*k*-Nearest Neighbour classifiers—A tutorial," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–25, Jul. 2022.
- [11] P. Fard Moshiri, R. Shahbazian, M. Nabati, and S. A. Ghorashi, "A CSI-based human activity recognition using deep learning," *Sensors*, vol. 21, no. 21, p. 7225, Oct. 2021.
- [12] D. Folgado, M. Barandas, M. Antunes, M. L. Nunes, H. Liu, Y. Hartmann, T. Schultz, and H. Gamboa, "TSSEARCH: Time series subsequence search library," *SoftwareX*, vol. 18, Jun. 2022, Art. no. 101049.
- [13] Y. Gu, L. Quan, and F. Ren, "Wi-Fi-assisted human activity recognition," in *Proc. IEEE Asia Pacific Conf. Wireless Mobile*, Aug. 2014, pp. 60–65.
- [14] Y. Gu, F. Ren, and J. Li, "PAWS: Passive human activity recognition based on Wi-Fi ambient signals," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 796–805, Oct. 2016.
- [15] Z. Guo, F. Xiao, B. Sheng, H. Fei, and S. Yu, "WiReader: Adaptive air handwriting recognition based on commercial Wi-Fi signal," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10483–10494, Oct. 2020.
- [16] Y. Hartmann, H. Liu, S. Lahrberg, and T. Schultz, "Interpretable high-level features for human activity recognition," in *Proc. 15th Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIOSIGNALS)*, 2022, pp. 40–49.
- [17] Y. Hartmann, H. Liu, and T. Schultz, "Feature space reduction for multimodal human activity recognition," in *Proc. 13th Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIOSIGNALS)*, 2020, pp. 135–140.
- [18] Y. Hartmann, H. Liu, and T. Schultz, "Interactive and interpretable online human activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Other Affiliated Events (PerCom Workshops)*, Mar. 2022, pp. 109–111.
- [19] Y. Hartmann, H. Liu, and T. Schultz, "High-level features for human activity recognition and modeling," in *Biomedical Engineering Systems and Technologies*, A. C. A. Roque, D. Gracanic, R. Lorenz, A. Tsanas, N. Bier, A. Fred, and H. Gamboa, Eds. Cham, Switzerland: Springer, 2023, pp. 141–163.
- [20] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10639–10651, Dec. 2019.
- [21] M. R. Hoque, M. H. Kabir, and S.-H. Yang, "Development of a cooperative middleware to provide context-aware service in smart home," *Int. J. Smart Home*, vol. 11, no. 5, pp. 33–40, May 2017.
- [22] M. S. Islam, M. K. A. Jannat, M. N. Hossain, W.-S. Kim, S.-W. Lee, and S.-H. Yang, "STC-NLSTMNet: An improved human activity recognition method using convolutional neural network with NLSTM from Wi-Fi CSI," *Sensors*, vol. 23, no. 1, p. 356, Dec. 2022.
- [23] M. S. Islam, K. Thapa, and S.-H. Yang, "Epileptic-Net: An improved epileptic seizure detection system using dense convolutional block with attention network from EEG," *Sensors*, vol. 22, no. 3, p. 728, Jan. 2022.
- [24] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 289–304.
- [25] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.
- [26] H. Li, X. He, X. Chen, Y. Fang, and Q. Fang, "Wi-Motion: A robust human activity recognition using Wi-Fi signals," *IEEE Access*, vol. 7, pp. 153287–153299, 2019.
- [27] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, Jan. 2019.
- [28] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [29] H. Liu, "Biosignal processing and activity modeling for multimodal human activity recognition," Ph.D. thesis, Fac. 3—Math. Comput. Sci., Universität Bremen, Bremen, Germany, 2021.
- [30] H. Liu, H. Gamboa, and T. Schultz, "Sensor-based human activity and behavior research: Where advanced sensing and recognition technologies meet," *Sensors*, vol. 23, no. 1, p. 125, Dec. 2022.
- [31] H. Liu, Y. Hartmann, and T. Schultz, "Motion units: Generalized sequence modeling of human activities for sensor-based activity recognition," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 1506–1510.
- [32] H. Liu and T. Schultz, "How long are various types of daily activities? Statistical analysis of a multimodal wearable sensor-based human activity dataset," in *Proc. 15th Int. Joint Conf. Biomed. Eng. Syst. Technol. (HEALTHINF)*, 2022, pp. 680–688.
- [33] H. Liu, T. Xue, and T. Schultz, "On a real real-time wearable human activity recognition system," in *Proc. 16th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2023, pp. 711–720.
- [34] S. N. Makhadmeh, A. T. Khader, M. A. Al-Betar, S. Naim, A. K. Abasi, and Z. A. A. Alyasseri, "Optimization methods for power scheduling problems in smart home: Survey," *Renew. Sustain. Energy Rev.*, vol. 115, Nov. 2019, Art. no. 109362.
- [35] S. A. E. Mohamed, "Why the accuracy of the received signal strengths as a positioning technique was not accurate?" *Int. J. Wireless Mobile Netw.*, vol. 3, no. 3, pp. 69–82, Jun. 2011.
- [36] P. F. Moshiri, H. Navidan, R. Shahbazian, S. A. Ghorashi, and D. Windridge, "Using GAN to enhance the accuracy of indoor human activity recognition," 2020, *arXiv:2004.11228*.
- [37] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–25, Jan. 2018.
- [38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [39] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2013, pp. 27–38.
- [40] J. Rodrigues, H. Liu, D. Folgado, D. Belo, T. Schultz, and H. Gamboa, "Feature-based information retrieval of multimodal biosignals with a self-similarity matrix: Focus on automatic segmentation," *Biosensors*, vol. 12, no. 12, p. 1182, Dec. 2022.
- [41] H. Salehinejad and S. Valaee, "LiteHAR: Lightweight human activity recognition from Wi-Fi signals with random convolution kernels," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 4068–4072.
- [42] E. Shalaby, N. ElShennawy, and A. Sarhan, "Utilizing deep learning models in CSI-based human activity recognition," *Neural Comput. Appl.*, vol. 34, no. 8, pp. 5993–6010, Apr. 2022.
- [43] S. Sigg, U. Blanke, and G. Tröster, "The telepathic phone: Frictionless activity recognition from Wi-Fi-RSSI," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2014, pp. 148–155.



- [44] S. Sigg, S. Shi, F. Buesching, Y. Ji, and L. Wolf, "Leveraging RF-channel fluctuation for activity recognition: Active and passive systems, continuous and RSSI-based signal features," in *Proc. Int. Conf. Adv. Mobile Comput. Multimedia (MoMM)*, 2013, pp. 43–52.
- [45] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
- [46] Y. Taso, S.-C. Yeh, Y.-Y. Liang, C.-H. Wang, and S.-H. Fang, "Subcarrier selection for efficient CSI-based indoor localization," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 383, Feb. 2018, Art. no. 012017.
- [47] M. H. Uddin, J. M. K. Ara, M. H. Rahman, and S. H. Yang, "A study of real-time physical activity recognition from motion sensors via smartphone using deep neural network," in *Proc. 5th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2021, pp. 1–6.
- [48] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2018, pp. 401–413.
- [49] F. Wang, S. Panev, Z. Dai, J. Han, and D. Huang, "Can WiFi estimate person pose?" 2019, *arXiv:1904.00277*.
- [50] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 593–604.
- [51] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.
- [52] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.
- [53] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.
- [54] T. Xue and H. Liu, "Hidden Markov model and its application in human activity recognition and fall detection: A review," in *Proc. 10th Int. Conf. Commun., Signal Process., Syst.*, vol. 1. Cham, Switzerland: Springer, 2022, pp. 863–869.
- [55] S. K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H. M. Pandey, and M. Mathur, "CSITime: Privacy-preserving human activity recognition using WiFi channel state information," *Neural Netw.*, vol. 146, pp. 11–21, Feb. 2022.
- [56] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.
- [57] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw.*, 2007, pp. 222–229.



**MIR KANON ARA JANNAT** received the B.Sc. degree in computer science and engineering from Islamic University, Kushtia, Bangladesh, in 2018, and the M.Sc. degree from the Department of Electronic Engineering, Kwangwoon University, Seoul, Republic of Korea, in 2023. Her current research interests include human activity recognition, digital signal processing, machine learning, and deep learning.



**MD. SHAFIQUK ISLAM** received the B.Sc. and M.Sc. degrees in computer science and engineering from Islamic University, Kushtia, Bangladesh, in 2016 and 2017, respectively, and the Ph.D. degree from the Department of Electronic Engineering, Kwangwoon University, Seoul, Republic of Korea, in 2023. His current research interests include human activity recognition, digital signal and image processing, and deep learning.



**SUNG-HYUN YANG** received the B.Sc. and M.Sc. degrees in electrical engineering and the Ph.D. degree from Kwangwoon University, Seoul, Republic of Korea, in 1983, 1987, and 1993 respectively. He was a Research Scientist with Boston University, from 1996 to 1998. He is currently a Professor of electronic engineering with Kwangwoon University. He is also the Director of the Ubiquitous Home Network Center, Kwangwoon University. His current research interests include digital logic, embedded systems, M2M, next-generation ubiquitous home networks, and context-aware systems. He was the Chairperson of the Home Network Market Activation Section, Korean Association for Smart Home, from 2007 to 2008.



**HUI LIU** (Member, IEEE) received the B.Sc. degree in information engineering from Shanghai Jiao Tong University (SJTU), in 2008, the Diplôme-Ingénieur degree in electrical engineering from the Technical University of Berlin, in 2009, the M.Sc. degree in communication and information systems from SJTU, in 2011, and the Ph.D. degree from the University of Bremen, in 2021. He has been a Researcher with the Cognitive Systems Laboratory, University of Bremen, since 2016. His current research interests include biosignal processing, human activity recognition, multimodal time-series analysis, and music information retrieval.

...