WiWeHAR: Multimodal Human Activity Recognition Using Wi-Fi and Wearable Sensing Modalities

MUHAMMAD MUAAZ [ID] [1], **ALI CHELLI** [1], **(Member, IEEE),**
AHMED ABDELMONEM ABDELGAWWAD [1], **ANDREU CATALÀ MALLOFRÉ** [2],
AND MATTHIAS PÄTZOLD [ID] [1], **(Senior Member, IEEE)**

[1]Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway
[2]Technical Research Centre for Dependency Care and Autonomous Living (CETpD), Universitat Politécnica de Catalunya–BarcelonaTech (UPC), 08800 Barcelona, Spain

Corresponding author: Muhammad Muaaz (muhammad.muaaz@uia.no)

ABSTRACT Robust and accurate human activity recognition (HAR) systems are essential to many human-centric services within active assisted living and healthcare facilities. Traditional HAR systems mostly leverage a single sensing modality (e.g., either wearable, vision, or radio frequency sensing) combined with machine learning techniques to recognize human activities. Such unimodal HAR systems do not cope well with real-time changes in the environment. To overcome this limitation, new HAR systems that incorporate multiple sensing modalities are needed. Multiple diverse sensors can provide more accurate and complete information resulting in better recognition of the performed activities. This article presents WiWeHAR—a multimodal HAR system that uses combined Wi-Fi and wearable sensing modalities to simultaneously sense the performed activities. WiWeHAR makes use of standard Wi-Fi network interface cards to collect the channel state information (CSI) and a wearable inertial measurement unit (IMU) consisting of accelerometer, gyroscope, magnetometer sensors to collect the user's local body movements. We compute the time-variant mean Doppler shift (MDS) from the processed CSI data and magnitude from the inertial data for each sensor of the IMU. Thereafter, we separately extract various time- and frequency-domain features from the magnitude data and the MDS. We apply feature-level fusion to combine the extracted features, and finally supervised learning techniques are used to recognize the performed activities. We evaluate the performance of WiWeHAR by using a multimodal human activity data set, which was obtained from 9 participants. Each participant carried out four activities, such as walking, falling, sitting, and picking up an object from the floor. Our results indicate that the proposed multimodal WiWeHAR system outperforms the unimodal CSI, accelerometer, gyroscope, and magnetometer HAR systems and achieves an overall recognition accuracy of 99.6%–100%.

INDEX TERMS Activity recognition, Doppler effect, feature extraction, feature fusion, machine learning, micro-Doppler signature, principal component analysis, radio frequency sensing, wearable sensing.

I. INTRODUCTION

Human activity recognition (HAR) commonly refers to automatically identifying physical activities performed by humans based on sensor data. HAR plays a crucial role in numerous human-centric applications that need to keep track of physical activities and behaviour of humans.

The associate editor coordinating the review of this manuscript and approving it for publication was Ting Wang [ID].

For instance, the aim of active assisted living (AAL) is to assist elderly and physically or cognitively impaired people to live independently in their homes by providing them with technical assistance in carrying out their daily activities [1], [2]. To provide such support, AAL systems proactively need to know the behaviour and the activities of the person. The information about the person's behaviour and activities can be obtained from a HAR system. Moreover, in geriatric clinical practice, various functionality tests are performed, such as the

timed up and go (TUG) [3] or more complete assessments that involve the quantification of the fundamental physiological, cognitive, and social parameters of comprehensive geriatric assessment (CGA) [4]. The analysis of the evolution of these parameters is used for the prescription of medication and/or preventive actions, with a clear impact on improving the quality of life of elderly. In general, the tests and measures described are carried out in medical examinations, but it would be very useful to automatically obtain information correlated with the assessment of functionality, i.e., by using systems that monitor daily life activities in a ubiquitous way. Furthermore, fall detection and prevention is essential for elderly healthcare and wellbeing. Because falls are important causes of different levels of trauma with fatal consequences in a high percentage; and psychological effects, such as the fear of falling, which leads to less activity implying a clear decline in functionality [5]. Therefore, health monitoring systems would profit from incorporating human activity trackers enabling to track activities of patients [6] and to detect accidental falls [7]. Upon detecting a fall or other needs of assistance, these healthcare systems can invoke a nearby hospital or an emergency care provider for assistance. Moreover, HAR systems are also used for indoor and outdoor surveillance and for detecting suspicious behaviour and activities in public places, such as airports [8], public transport [9], and correctional facilities [10].

Typically, the architecture of HAR systems consist of the following modules: (1) a sensing module—to collect the sensor data while a person is performing the activities; (2) a data processing module—to process the raw sensor data to eliminate noise from the data, and to normalize the data; (3) a segmentation module—to identify and extract the segments of interest (e.g., the time period when the person is performing an activity) from the processed data; (4) a feature extraction module—to extract features from the segmented data; and finally (5) a classification module that applies machine learning or deep learning techniques to predict the activities performed by the person. Moreover, HAR systems are generally divided into device-based, device-free, and multimodal HAR systems based on the underlying sensor technology that is used for data collection.

A. DEVICE-BASED HAR SYSTEMS

Device-based
HAR systems, also known as wearable sensor-based HAR systems, require the user to either wear or carry sensors, such as accelerometers and gyroscopes. The data collected by these sensors are used to recognize human activities. Smartphones [11]–[13], smartwatches [14], [15], and dedicated inertial measurement units (IMUs) [16] are commonly used wearable sensor platforms for device-based HAR systems, gestures recognition systems, and health monitoring systems. In recent years, smart clothes with integrated sensors have been designed to monitor vital signs and human activities. For instance, in [17], a textile-based wearable sensing platform has been presented that consists of a sensing shirt,

a sensing trouser, sensing gloves, and sensing shoes. The sensing modules of this wearable sensing platform are embedded with inertial sensors, textile electromyography (EMG) electrodes, and piezoresistive sensors to collect data that are used to monitor the physical activity of stroke survivors and evaluate their recovery. Similarly, Hexoskin [18] offers clinically validated and machine washable smart garments, e.g., smart shirts for continuous monitoring of heart and breathing rate, and activity data that can effectively be used to recognize human activities.

Although device-based HAR systems can recognize human activities with great accuracy, they have several drawbacks. For example, the wearable sensors must be worn all the time for continuous activity recognition, which might be invasive, less comfortable, and annoying for cognitively impaired individuals and elderly. Moreover, the wearable sensors must be placed on the human body correctly according to the guidelines of the sensor providers. Otherwise, wrong placement of the sensors may provide inaccurate measurements that can lead to incorrect results or reduced recognition accuracy [19]–[21].

B. DEVICE-FREE HAR SYSTEMS

In device-free HAR systems, vision and radio frequency (RF) sensors are commonly leveraged to identify human activities. Vision-based HAR systems are widely used for security and surveillance applications. They use cameras to record videos or capture still images of individuals while users perform activities. Thereafter, computer vision techniques are applied to classify the activities [22]. Vision-based HAR systems are generally very accurate, and they can effectively track multiple persons, but they suffer from various limitations. For instance, they typically have a limited monitoring area; they need a clear view of the monitoring area; their performance highly depends on the lighting conditions; and the recorded videos might raise concerns about user privacy.

RF-based device-free HAR systems rely on the phenomenon that human bodies reflect RF signals, and movements of different human body segments cause variations in the frequencies of the RF signals due to the Doppler effect. These RF signals, enriched with Doppler frequency variations, are collected and used for human activity recognition [23], [24], gesture recognition [25], and fall detection [26], [27]. RF-based HAR systems are generally divided into standard Wi-Fi-based [23], [24], [28] and radar-based [29], [30] systems. Wi-Fi-based HAR systems are further divided into two categories. The first category uses the received signal strength indicator (RSSI) attribute of Wi-Fi signals [24], [31], which represents the variation of the received signal strength during propagation. Whereas, the second category exploits the channel state information (CSI) [23], [32], [33], which represents the amplitude and phase information associated with each orthogonal frequency-division multiplexing (OFDM) subcarrier. Therefore, the CSI is considered finer grained, and it has been shown that CSI-based HAR systems perform better than RSSI-based HAR systems [23].

Wi-Fi- and radar-based HAR systems have their own advantages and disadvantages. For example, a Wi-Fi-based HAR system can effectively make use of ubiquitous Wi-Fi infrastructure or off-the-shelf Wi-Fi access points to collect the CSI, which can be used to recognize human activities. Therefore, Wi-Fi-based HAR systems are highly cost-effective. Radar-based systems offer better spatial resolution compared to Wi-Fi-based system, but this advantage comes with significantly higher costs. Consequently, radar-based HAR systems have better recognition accuracy, but the high costs limit their widespread use.

C. MULTIMODAL HAR SYSTEMS

Traditional HAR systems mostly employ a single sensing modality, such as either an RF-based or a vision-based sensing modality. However, a single sensing modality cannot handle the changing conditions of the environment in real time. For instance, vision- or RF-based HAR systems can only recognize human activities within the monitoring area. Therefore, multimodal HAR systems are used to overcome the limitation of unimodal HAR systems. In multimodal HAR systems, two or more sensing modalities are simultaneously used to collect the data while a person is performing activities. For instance, wearable and RF sensing modalities offer complementary information, because the mean Doppler shift (MDS) computed from the Wi-Fi CSI provides a general view of motion, whereas the wearable inertial sensor provides information about the local body movements. By merging the data of these two complementary sensor modalities, a robust HAR system can be designed. The fusion of the collected data can be performed either at the data level, the feature level, or the decision level [34]. Data-level fusion aims to generate new raw data by combining the raw data obtained from different homogeneous sensing modalities, where the new raw data are generally considered to be more informative compared to the raw data obtained from a single sensor. Feature-level fusion combines the features extracted from the processed data of each sensing modality. Finally, decision-level fusion generates a single decision by combining the decisions provided by the unimodal classifiers.

The literature on multimodal HAR techniques is quite diverse due to a large number of available sensing modalities. For instance, in [35], the decision-level fusion technique was used to combine the decisions of vision- and Wi-Fi-based sensing modalities, aiming to improve the classification results of common human activities, such as walking, sitting, and standing. In [36], an ensemble learning technique was used to combine the results of individual deep learning classifiers that were trained and evaluated using the data collected with multiple wearable inertial sensors. The results in [36] show that this multimodal wearable sensing approach leads to improved recognition accuracy. In [37], Wi-Fi CSI and accelerometer sensors were used to collect human motion data while playing table tennis strokes, such as forehand lobe and drive, backhand lob and twist, hop step, step, and squat. First, for each sensing modality, they used a unimodal hidden

Markov model (HMM) to recognize the table tennis motions. Later, decision level fusion was used to combine the decisions of both unimodal HMMs. The proposed approach in [37] can recognize table tennis motions with an overall accuracy of 97%. Sensing-Fi is a multimodal fall detection system that incorporates Wi-Fi CSI and acceleration data recorded with a ground-mounted accelerometer [38]. Sensing-Fi detects falls with 95% accuracy by employing the decision-level fusion that combines the decisions of individual sensing modalities.

This article presents WiWeHAR—a multimodal HAR system that recognizes human activities by fusing the complimentary Wi-Fi and wearable sensing modalities. The Wi-Fi sensing modality captures the CSI using standard Wi-Fi network interface cards (NICs), and the local body movements are measured using a wearable IMU that consists of an accelerometer, a gyroscope, and a magnetometer sensor. We use the spectrogram method to estimate the MDS from the CSI data, whereas for each sensor of the IMU, the magnitude is computed from the corresponding sensor data. The contributions of this work are as follow: (1) we collect multimodal Wi-Fi CSI and inertial data from nine participants while carrying out four different activities by using standard NICs and an IMU in an indoor environment; (2) we apply the feature-level fusion technique to combine the features extracted from the magnitude data of each sensor of the IMU and the MDS, which is novel in itself; (3) we evaluate the performance of the proposed WiWeHAR system by using a support vector machine (SVM); and (4) we compare the performance of WiWeHAR with unimodal CSI- and unimodal IMU-based HAR systems that are developed in this work in terms of precision, recall, and overall accuracy.

The experimental design of this work is oriented towards the identification of activities including walking, sitting, falling, and picking up an object from the floor. The underlying idea is to demonstrate the reliability of the WiWeHAR system for subsequent use in more precise studies and close to clinical practice as may be the realization of the systematic TUG test. In the TUG test, it is only necessary to identify the time spent for sitting and getting up from a chair and the time spent for walking 3 m and making a 360-degree turn. The second target of WiWeHAR is the precise detection of falls with the possibility to link fall incidents to an emergency system. With the proposed system and the results obtained, the identification and use of daily life activities for therapeutic purposes is increasingly close and coincides with the recommendations of global public health agencies [39]. Moreover, the proposed WiWeHAR system systematically demonstrates that the Wi-Fi CSI and wearable sensing modalities are complementary to each other because their fusion significantly improves the recognition of human activities. Therefore, the proposed WiWeHAR system is suitable for indoor applications requiring a high degree of reliability in the accurate classification of human activities in indoor environments.

The rest of the article is organized as follows. A succinct overview of WiWeHAR is given in Section II, which is

followed by a comprehensive review of RF channel modelling for HAR systems in Section III. We explain the layout of our experimental setup and discuss the data collection process in Section IV. The steps involved in pre-processing and segmenting the CSI and inertial data are presented in Section V. The steps required to classify the data (i.e., feature extraction, feature fusion, performance metrics) and the classification results are presented in Section VI. Finally, in Section VII, we present the conclusion of this work and give directions for future work.

II. OVERVIEW OF THE WiWeHAR SYSTEM

The WiWeHAR system consists of multimodal sensing and machine learning phases. In the sensing phase, WiWeHAR uses RF and wearable sensing modalities to capture the CSI and the inertial data simultaneously, while a user is performing different activities as shown in Fig. 1. To collect the inertial data, we attach a wearable IMU at the lower back of the user. The IMU consists of a tri-axis accelerometer, a tri-axis gyroscope, and a tri-axis magnetometer sensor.
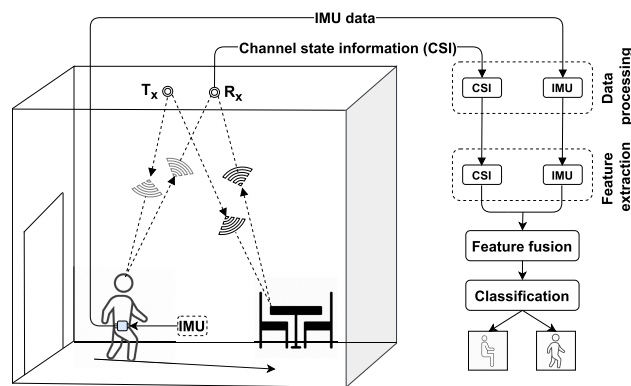

FIGURE 1. Overview of the proposed multimodal Wi-Fi- and wearable sensor-based HAR system.

To capture the CSI data, a Wi-Fi transmitter (T_x) and a Wi-Fi receiver (R_x) are deployed in the indoor environment. The T_x and R_x are standard Wi-Fi NICs [40] that are configured to operate in the 5 GHz Wi-Fi band. Wi-Fi in 5 GHz band has 45 channels and to avoid interference to and from other wireless networks in the vicinity, the WiWeHAR system can be configured to use a channel that is not used for the communication purpose. The T_x emits electromagnetic waves that propagate in the indoor environment. On their way from the T_x to the R_x, these electromagnetic waves reflect off the static objects (e.g., ceiling, furniture, and walls) and dynamic objects (e.g., body segments of a moving person). Moreover, the movements of the person in the indoor environment introduce variations in the frequencies of ambient electromagnetic waves. This phenomenon is called the Doppler effect. The Doppler effect is movement dependent, i.e., different types of activities performed by a person lead to distinct Doppler shift patterns.

First, we process the CSI and inertial data to reduce the impact of noise. The processed CSI data are used to compute the spectrogram that shows the time-variant Doppler characteristics of the radio channel caused by moving scatterers (e.g., moving person) and static (e.g., walls, furniture) scatterers. The variation in the spectral components is merely due to moving body segments in an otherwise static environment. Hence, the performance of the WiWeHAR system will be unaffected by the different placements of the static objects in the environment. In the subsequent step, we use the spectrogram method to estimate the MDS whereas, the collected inertial data is used to compute the magnitude for each sensor of the IMU. In the machine learning phase, we extract various time- and frequency-domain features from the MDS and magnitude data of each sensor of the IMU. These features are fused together and used to train a classifier to predict the labels of the activities performed by a user.

III. CHANNEL MODELING BACKGROUND

As mentioned in Section II, a Wi-Fi T_x and a Wi-Fi R_x are deployed in the indoor environment to collect the CSI. The T_x and R_x are configured to operate in the Wi-Fi injector-monitor mode. The T_x injects random data packets in the wireless channel, and the R_x estimates and reports the CSI along 30 OFDM subcarriers for each received data packet. The collected CSI data are represented in the form of a matrix with dimensions $N_{T_x} \times N_{R_x} \times K$, where N_{T_x}, N_{R_x}, and K indicate the number of antennas at the T_x side, the number of antennas at the receiver R_x side, and the number of OFDM subcarriers, respectively. In our case, we used a single transmit antenna and two receive antennas, i.e., $N_{T_x} = 1$ and $N_{R_x} = 2$, which implies that there exist two transmission links. Within the scope of this work, a transmission link is defined as the pair of a transmit and receive antenna. Therefore, our CSI data matrix has the dimensions $1 \times 2 \times 30$. The time-series of the CSI values for a given transmission link and subcarrier is known as a CSI stream. A comprehensive discussion about our experimental setup for CSI measurements is given in Section IV.

The measured CSI can be regarded as the channel transfer function (CTF)[1] $H_{i,j}(f'_k, t)$ of a transmission link indicating the link between the ith transmit antenna and the jth receive antenna sampled at the kth subcarrier f'_k. The expression of f'_k is given as

$$f'_k = f'_0 + k \Delta f' \quad (1)$$

where f'_0, k, and $\Delta f'$ stand for the carrier frequency, the subcarrier index, and the subcarrier bandwidth, respectively.

The influence of human activities on the indoor wireless channel can be explained with the help of a channel model, which collectively captures the contributions of moving and

---

[1]Within the context of this work, the terms CSI and CTF are used interchangeably.

fixed scatterers to the CTF[2] $H(f'_k, t)$. In our measurement campaign, we have ensured that only a single person is moving, and all of the other objects are static inside the room.

Usually, there exist a variety of static objects (e.g., walls, furniture, and ceiling) in a room. These static objects can be modelled by N fixed scatterers S^F_n ($n = 1, 2, \ldots, N$). Although we have considered that a single person is moving in the room, a single person in motion can be considered as a cluster of different moving body segments (e.g., head, torso, arms, and legs). Therefore, we model the moving person by a cluster of M moving scatterers S^M_m ($m = 1, 2, \ldots, M$). We can present the CTF of our indoor wireless channel model by summing the terms associated with N fixed and M moving scatterers as [41]

$$H(f'_k, t) = \sum_{n=1}^{N} c_n \exp\left(j\theta_{n,k}\right) + \sum_{m=1}^{M} c_m \exp\left(j\theta_{m,k}(t)\right) \quad (2)$$

where c_n indicates the path gain of the nth fixed scatterer. Similarly, the symbol c_m represents the path gain associated with the mth moving scatterer. The phase associated with the nth fixed scatterer and the kth subcarrier is given by $\theta_{n,k}$. Similarly, $\theta_{m,k}(t)$ indicates the phase of the mth moving scatterer and the kth subcarrier. Note that the phase $\theta_{m,k}(t)$ associated with the moving scatterer is time variant, while the phases $\theta_{n,k}$ associated with the fixed scatterers are random variables, which are modeled as independent identically distributed (i.i.d.) random variables, each of which is uniformly distributed in the interval $[0, 2\pi)$.

As we know, only the moving scatterers cause the Doppler shift in the RF signals, but not the fixed scatterers. We can compute the Doppler shift associated with the mth moving scatterer by using the following expression [41]

$$f_m(t) = -f_{m,\max}(t) \left\{ \cos\left(\beta^v_m(t)\right) \left[\cos\left(\beta^T_m(t)\right) \right. \right.$$
$$\cos\left(\alpha^T_m(t) - \alpha^v_m(t)\right)$$
$$\left. + \cos\left(\beta^R_m(t)\right) \cos\left(\alpha^v_m(t) - \alpha^R_m(t)\right) \right]$$
$$\left. + \sin\left(\beta^v_m(t)\right) \left[\sin\left(\beta^T_m(t)\right) + \sin\left(\beta^R_m(t)\right) \right] \right\} \quad (3)$$

where $\alpha^v_m(t)$ and $\beta^v_m(t)$ denote the azimuth and elevation angles of motion, respectively. The elevation angles of arrival and departure are indicated by $\beta^R_m(t)$ and $\beta^T_m(t)$, respectively. The azimuth angles of arrival and departure are represented by $\alpha^R_m(t)$, and $\alpha^T_m(t)$, respectively. Moreover, $f_{m,\max}(t)$ designates the maximum Doppler frequency, which is given as

$$f_{m,\max}(t) = \frac{v_m(t)}{c_0} f'_k \quad (4)$$

where $v_m(t)$ and c_0 indicate the time-variant speed of the mth moving scatterer and the speed of light, respectively.

---

[2]In the remaining part of this subsection, we do not use the subscripts that indicate the transmit and receive antenna indices (i.e., i and j) to simplify the notations.

We can represent the phase $\theta_{m,k}(t)$ of the mth moving scatterer in terms of the Doppler frequency as [42]

$$\theta_{m,k}(t) = 2\pi \int_{-\infty}^{t} f_m(u)du = \theta_{m,k} + 2\pi \int_{0}^{t} f_m(u)du \quad (5)$$

where the initial phase shift at time $t = 0$ is indicated by $\theta_{m,k}$, which is modelled by a uniformly distributed random variable in the interval $[0, 2\pi)$. As mentioned earlier in this section, the CSI streams correspond to the CTF $H(f'_k, t)$ values associated with a specific subchannel at subcarrier f'_k. Therefore, we can argue that every CSI data stream can be described by the above channel model presented in (2).

IV. EXPERIMENTAL SETUP AND DATA COLLECTION

We used two laptops and an IMU to simultaneously collect the CSI and inertial data. To collect and parse the CSI data, we installed the CSI tool [43] on both laptops. Each of the laptops was equipped with an Intel 5300 Wi-Fi NIC, which was configured to operate at 5.745 GHz central frequency in single-input multiple-output (SIMO) transmission mode with a bandwidth of 20 MHz. We used the injector-monitor Wi-Fi mode where one laptop served as a Wi-Fi T_x and the other as a Wi-Fi R_x. As the internal wireless antennas of the laptops have small antenna gains, we used external wireless antennas that were attached to the laptops using 141-1MSM+ RF cables. On the T_x side, we used a single directional antenna, whereas a directional and an omnidirectional antenna have been used on the R_x side. These antennas were mounted on a table of 0.8 m height, as shown in Fig. 2. During the data collection, the T_x was set to inject random data packets into the wireless channel at a sampling rate of 1 kHz, and the R_x captured the injected packets. For each received data packet, the CSI tool collected the CSI data of 30 OFDM subcarriers per transmission link.
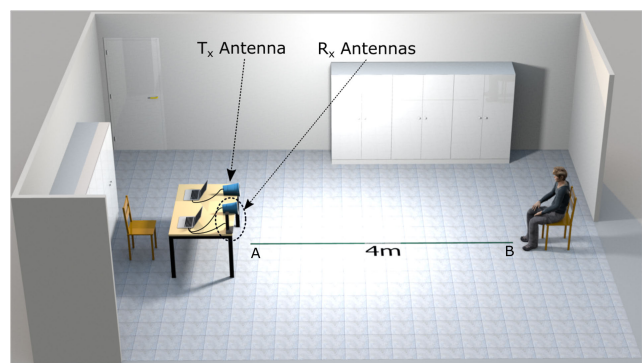

FIGURE 2. Experimental setup for data collection.

To collect the inertial data, we attached an IMU to the participants' lower back (the lumbar portion of the spine) as shown in Fig. 3. The IMU contains a tri-axis accelerometer, a tri-axis gyroscope, and a tri-axis magnetometer. Each sensor of the IMU was set to collect data at a sampling rate of 400 Hz. The IMU locally stored the raw data of each sensor, which were later transferred to the computer for further processing.

The human activity data were collected from nine healthy participants. These participants agreed to participate in the study by signing an informed consent. The demographics of the volunteers of this study are given in Table 1. Each participant performed four activities: walking, sitting on a chair, falling on a mattress, and picking up a small object from the floor. As the Doppler shift is sensitive to movements in the environment, we ensured that only a single person was carrying out activities during the data collection.

TABLE 1. Age, gender, weight, and height of the participants of the study.

Participant	Age (years)	Gender (male/female)	Weight (kg)	Height (cm)
1	22	m	76	180
2	25	m	77	178
3	63	m	73	170
4	39	f	66	166
5	42	m	90	190
6	40	f	53.5	165
7	23	m	78	180
8	22	m	76	180
9	25	m	77	178

The protocol of the experiment was as follows. The participants were asked to remain still for one second before starting and after completing an activity. Every participant performed the walking activity 10 times by walking back and forth from Point A to Point B (see Fig. 2), as they would walk in their routine life. For the falling activity, a 200 cm long, 90 cm wide, and 15 cm high mattress was placed at Point B, and the participants were told to stand close to the short side of the mattress and then fall onto the mattress. Each participant carried out the falling activity 10 times, out of which five times they fell on the mattress facing towards the antennas and five times facing away from the antennas. For the sitting activity, we placed a chair at Point B and asked the participants to stand still next to the chair facing towards the antennas and then sit on the chair as they would sit on an chair in their daily life. Each participant carried out the sitting activity five times. For the last activity, a small object was placed on the floor at Point B. The participants were asked to

stand close to the object and then pick up the object. Every participant conducted the picking activity five times. Three out of the four activities (sitting, falling, and picking up an object) were carried out at a distance of 4 m from the T_x and R_x antennas.

V. DATA PROCESSING
A. CSI DATA PROCESSING
The raw CSI data obtained from the Intel 5300 Wi-Fi NIC was very noisy. Therefore, the raw CSI data cannot directly be used to estimate the MDS caused by the moving person. As we know from (2), the CSI data consist of amplitude and phase information. The amplitude of the CSI data is affected by the burst noise and high amplitude impulse that are caused by the adaptive change in the transmission power and transmission rate [23], [44], [45]. On the other hand, the phase of the CSI data is distorted by errors such as carrier frequency offset (CFO) and sampling frequency offset (SFO) [44], [46]. These errors mainly occur due to asynchronous clocks of the T_x and R_x [44]–[46].

1) PHASE CORRECTION
The first step towards estimating the MDS requires the elimination of the CSI phase distortions. There are several methods that can be employed to eliminate phase distortions. For example, the phase sanitization method [25], [47], the phase calibration method [46], [48], and the CSI ratio method [49]. In the phase sanitization method, a linear transformation is applied to determine the true phases from the measured distorted phases of 30 OFDM subcarriers. Although the transformed phases show some patterns compared to the measured phases, another study [48] reported that the transformed phases still do not accurately disclose the Doppler characteristics of the measured CSI data.

The second phase correction approach is called the back-to-back phase calibration method. This method requires additional hardware in the form of a two-way splitter to establish a back-to-back channel connection [46]. One of the RF antenna ports of the Wi-Fi NIC of the T_x is connected to the input of the splitter. One of the two outputs of the splitter is connected to the T_x antenna, whereas the other output is connected to one of the RF antenna ports of the Wi-Fi NIC at the R_x. The splitter splits the transmitted RF signal into two identical signals, where one output signal is transmitted via the T_x antenna and the other output signal is transmitted via the back-to-back channel. Since the data received over the back-to-back channel do not suffer from phase distortions, they can be used to correct the phase of the data received by the R_x antenna.

In this work, we employed the CSI ratio method [49] to reduce phase distortions. In contrast to the back-to-back phase calibration technique, the CSI ratio method does not require additional hardware. In the CSI ratio method, two receiving antennas are connected to the same receiver for simultaneous collection of the CSI data $H_{1,1}(f'_k, t)$ and $H_{1,2}(f'_k, t)$. Thereafter, the CSI ratio is computed by

dividing the CSI data obtained from the first and the second R_x antenna, i.e.,

$$R(f'_k, t) = \frac{H_{1,1}(f'_k, t)}{H_{1,2}(f'_k, t)}. \tag{6}$$

It has been shown that this element-wise division operation removes most of the noise from the amplitude and phase information [49]. In our experiments, we also observed that the spectrogram obtained after applying the CSI ratio method shows clear micro-Doppler signatures of the performed activity, compared to the spectrogram obtained from the CSI data of any of the individual transmission links. A comparison of the spectrograms that are estimated with and without using the CSI ratio method is given in Fig. 4. Note that, we used the data of the same activity trial carried out by the same participant to compute these spectrograms. Moreover, all remaining data processing steps explained in Subsections V-A2 and V-A3 were left unchanged.
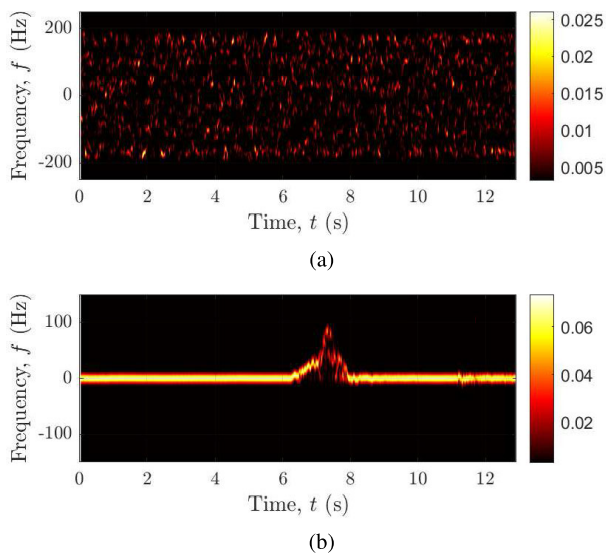

FIGURE 4. A comparison of spectrograms of the falling activity obtained (a) without and (b) with applying the CSI ratio method.

2) DIMENSIONALITY AND NOISE REDUCTION

As mentioned in Section IV, the CSI tool collects a total of 30 CSI data streams for each transmission link. Therefore, after computing the CSI ratio, we obtain 30 CSI ratio streams $R(f'_k, t)$. Previous work has shown that the variations of RF signals caused by human body movements are correlated across different CSI streams [23]. Therefore, the principal component analysis (PCA) [50] was used to remove the correlated information from the CSI ratio streams.

The PCA is usually applied to real-valued data sets. In our case, however, the CSI ratio streams $R(f'_k, t)$ are complex-valued. Therefore, we used the adaptation of the PCA in [51] for the complex-valued data, which directly extends the original PCA to the complex domain in a straightforward way. To apply the complex PCA, we first arrange the CSI ratio

streams in the form of a matrix. To do so, we consider the samples of $R(f'_k, t)$ at $t = t_n = nT$ for $n = 1, 2, \ldots, N$, where N denotes the number of samples in the time domain, and T is the sampling interval. This leads to $R_{kn} = R(f'_k, t_n)$, which allows us to define the CSI ratio matrix

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1} & R_{K2} & \cdots & R_{KN} \end{pmatrix} \in \mathbb{C}^{K \times N}. \tag{7}$$

In the next step, the CSI ratio matrix $\mathbf{R}$ is centered by subtracting the sample mean value $\bar{m}_k$, defined as $\bar{m}_k = \sum_{n=1}^{N} R_{kn}/N$, from each CSI ratio stream, i.e.,

$$\mathbf{R}_c = \begin{pmatrix} R_{11} - \bar{m}_1 & R_{12} - \bar{m}_1 & \cdots & R_{1N} - \bar{m}_1 \\ R_{21} - \bar{m}_2 & R_{22} - \bar{m}_2 & \cdots & R_{2N} - \bar{m}_2 \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1} - \bar{m}_K & R_{K2} - \bar{m}_K & \cdots & R_{KN} - \bar{m}_K \end{pmatrix} \in \mathbb{C}^{K \times N}. \tag{8}$$

The centered CSI ratio matrix is used to compute the covariance matrix $\mathbf{C}$ according to

$$\mathbf{C} = \frac{1}{N-1} \mathbf{R}_c \mathbf{R}_c^{\mathbf{H}} \tag{9}$$

where $\mathbf{R}_c^{\mathbf{H}}$ indicates the Hermitian (or conjugate transpose) of the centered CSI ratio matrix $\mathbf{R}_c$. Thereafter, the covariance matrix $\mathbf{C}$ is factorized into eigenvectors and eigenvalues.

As the covariance matrix $\mathbf{C}$ is a Hermitian matrix of dimension $K \times K$, its eigendecomposition can result in at most K real-valued eigenvalues and K complex-valued eigenvectors. Let $\mathbf{Z}$ be the matrix consisting of K eigenvectors, which are arranged such that the first eigenvector corresponds to the highest eigenvalue, and the second eigenvector corresponds to the second highest eigenvalue, etc. These eigenvectors are called the principal axes. To obtain the principal components, we project the centered CSI ratio matrix $\mathbf{R}_c$ onto these principal axes according to

$$\mathbf{Y} = \mathbf{R}_c^{\mathsf{T}} \mathbf{Z} \tag{10}$$

where $(.)^{\mathsf{T}}$ denotes the transpose operator. The first principal component in $\mathbf{Y}$ has the highest possible variance and captures the maximum information, whereas the last principal component in $\mathbf{Y}$ has the lowest possible variance and captures the minimum information of the original data. The amplitude plots of the first six principal components obtained from the CSI ratio data of the falling activity are shown in Fig. 5. In Fig. 5, the first two principal components (Fig. 5(a) and Fig. 5(b)) show the variation caused by the human movement, where the first principal component is least affected by noise. The next four principal components (Figs. 5(c) − 5(f)) show mainly the random noise present in the data. Moreover, we can observe that each succeeding principal component is noisier than its preceding one. Therefore, we used only the first principal component for the subsequent steps.
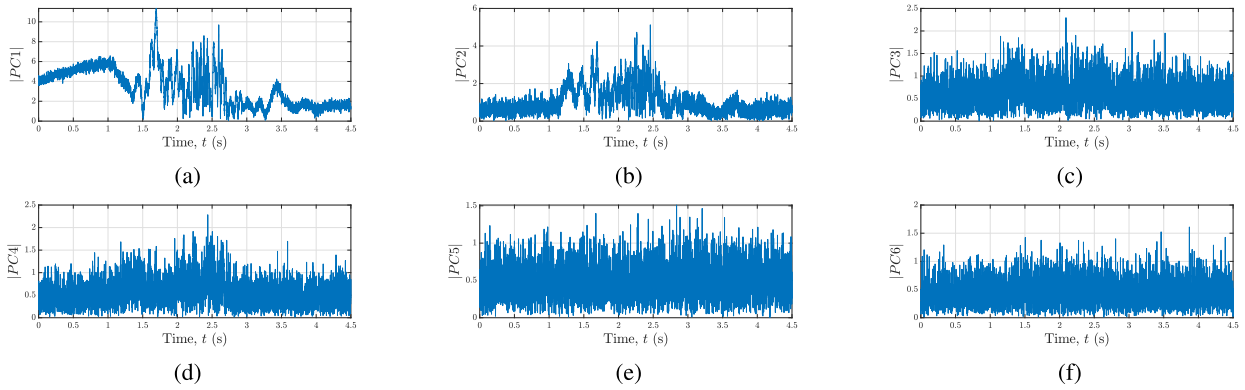
FIGURE 5. The amplitude plots of the first six principal components of the falling activity.

The advantage of the PCA is threefold. Firstly, by using the first principal component, the 30 CSI ratio data streams are reduced to a single data stream, which significantly reduces the overall computational load of our approach. Secondly, it significantly reduces the residual noise present in the measurement data. Thirdly, the effect of the fixed scatterers is also greatly reduced without applying a high pass filter as we can see in the spectrograms shown in Fig. 6. Furthermore, a low pass filter is used to further minimize the impact of high frequency components (that are not caused by human activities) from the selected principal component data. We set the cutoff frequency of the low pass filter to 150 Hz as the maximum Doppler shift can reach up to 110 Hz for the falling activity (see Fig. 6(a)). For the other three activities, the maximum Doppler shift is less than 110 Hz (see Fig. 6(c), Fig. 6(e), and Fig. 6(g)).

3) ESTIMATING THE SPECTROGRAM AND THE MDS
We apply the spectrogram technique to estimate the time-variant Doppler power spectrum of the filtered principal component. First, we multiply the filtered data with a sliding Gaussian window

$$g(t) = \frac{1}{\sqrt{\sigma_w \sqrt{\pi}}} e^{-\frac{t^2}{2\sigma_w^2}} \qquad (11)$$

which is positive, even, and has unit energy. In (11), σ_w indicates the spread of the Gaussian window. As the window slides over the filtered data with respect to time, we compute the short-time Fourier transform (STFT) of the windowed filtered data according to

$$X_{Y_1}(f, t) = \int_{-\infty}^{\infty} Y_1(f, t')g(t' - t)e^{-j2\pi ft'} dt' \qquad (12)$$

where $Y_1(f, t')$ denotes the first principal component; and symbols t' and t indicate the running time and the local time, respectively. In the next step, the spectrogram is computed by squaring the magnitude of the STFT, i.e., [52]

$$S_{Y_1}(f, t) = |X_{Y_1}(f, t)|^2. \qquad (13)$$

Finally, we estimate the time-variant MDS $B_{Y_1}^{(1)}(t)$ from the spectrogram by using

$$B_{Y_1}^{(1)}(t) = \frac{\int_{-\infty}^{\infty} fS_{Y_1}(f, t)df}{\int_{-\infty}^{\infty} S_{Y_1}(f, t)df}. \qquad (14)$$

The spectrograms and the corresponding MDSs of the four activities are shown in Fig. 6.

B. IMU DATA PROCESSING
Unlike CSI data processing, fewer steps are required to process the inertial data. The IMU used for data collection consists of a tri-axis accelerometer, a tri-axis gyroscope, and a tri-axis magnetometer. Thus, the inertial data include three-dimensional accelerometer data, three-dimensional gyroscope data, and three-dimensional magnetometer data. Recall that during the data collection process, the IMU was attached to the lower back of each participant. As a result, the orientation of the IMU might have slightly changed while the participant was performing different activities. Therefore, to mitigate the effect of the changes in the orientation of the IMU, we compute the magnitude $mag(t)$ for each sensor type in the IMU using its three-dimensional data, which is defined as

$$mag(t) = \sqrt{x^2(t) + y^2(t) + z^2(t)} \qquad (15)$$

where $x(t)$, $y(t)$, and $z(t)$ indicate the raw sensor data recorded along x, y, and z direction. To further minimize the impact of noise from the magnitude of each sensor, we apply the locally estimated scatterplot smoothing (LOESS) filter with a window size of 125 ms.

C. ACTIVE SEGMENT EXTRACTION
Recall that the CSI and inertial data were recorded simultaneously. As the participants were asked to remain inactive for at least one second before starting and after completing an activity. Therefore, we do not observe significant variations in the magnitude and MDS at the beginning and the end of
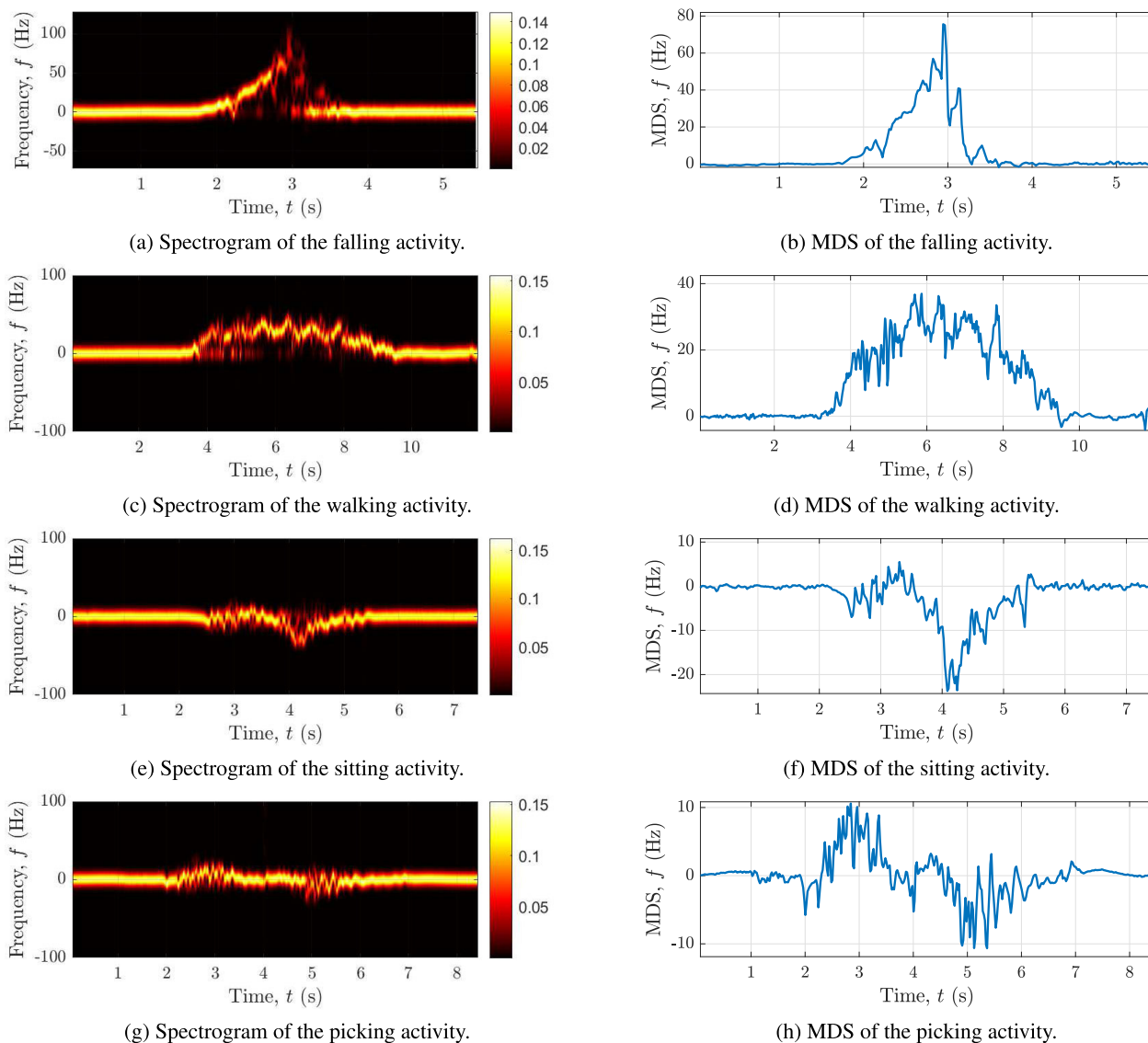
(a) Spectrogram of the falling activity.

(b) MDS of the falling activity.

(c) Spectrogram of the walking activity.

(d) MDS of the walking activity.

(e) Spectrogram of the sitting activity.

(f) MDS of the sitting activity.

(g) Spectrogram of the picking activity.

(h) MDS of the picking activity.

FIGURE 6. The spectrograms and their corresponding MDSs of the four activities: falling, walking, sitting, and picking up an object from the floor.

each activity, as shown in Fig. 6. We apply the variance-based thresholding method (VTM) [53] to automatically determine the beginning and end of the active segments from the estimated MDS and the magnitude data of accelerometer, gyroscope, and magnetometer. Within the context of this work, an active segment indicates the time interval of the MDS or the magnitude within which the person is active. To extract an active segment from the MDS, the variance of the MDS is monitored over a rectangular sliding time window of size 0.1 s. As the window slides over time, we compute the variance of the windowed-MDS and compare it to a certain threshold value, where a crossing of the variance above or below the predefined threshold value indicates the start or the end of an active MDS segment, respectively. Once the start and end of all the active MDS segments are marked, we pick a segment that is longer than 1.5 s. This is done to discard the

insignificant changes in the MDS that might have occurred due to slight movements of body segments when the subject is not performing the activity. Fig. 7 shows the start and end of the active MDS segment determined by using the VTM.
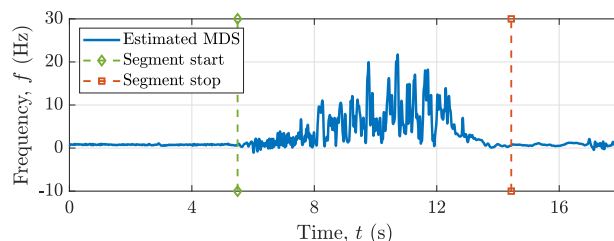

FIGURE 7. Extracting an active MDS segment from the estimated MDS.

Similarly, we extract an active magnitude segments from the magnitude data of each sensor of the IMU by adapting the appropriate window size and variance threshold.

VI. CLASSIFYING HUMAN ACTIVITIES

A. FEATURE EXTRACTION AND FEATURE FUSION

In machine learning parlance, feature extraction is a preliminary step to building an effective classifier. Usually, the features are selected directly from raw or processed data, or they are derived from raw or processed data. Generally, these features are uncorrelated variables that are fed to the classifier. The main reasons for feature extraction are twofold. First, reducing the number of variables of the data because classifiers trained with a large number of variables do not generalize well, and consequently, their performance in evaluating against unseen data is significantly reduced. Second, reducing the resources required to build the classifiers in terms of computational power, memory, and time consumption. As mentioned in Section IV, the participants performed four different activities, where each activity was repeated several times. Within the scope of this work, each repetition of an activity is called the activity sample. Although we recorded the CSI and inertial data simultaneously, these data were processed separately. First, for each activity sample, we extracted an active MDS segment from the CSI data and active magnitude segments from the magnitude data of each sensor of the IMU. Then, these active segments were labelled according to the type of activity.

Thereafter, we extracted a CSI-only, an accelerometer-only, a gyroscope-only, and a magnetometer-only feature vector from the active MDS segment, the active acceleration magnitude segment, the active gyroscope magnitude segment, and the active magnetometer magnitude segment, respectively. The CSI-only, accelerometer-only, and the gyroscope-only feature vectors consist of the same 23 time- and frequency-domain features, whereas the magnetometer-only feature vector consists of the first 22 time- and frequency-domain features described in the Appendix. The CSI-only feature vectors extracted from all activity samples were combined to form a CSI-only feature set. Analogously, we prepared the accelerometer-only, gyroscope-only, and magnetometer-only features sets. Within the scope of this work, these feature sets are called unimodal feature sets, as they are separately extracted for each sensor from its data.

Thereafter, we used the min-max normalization technique [54] to normalize each feature set according to

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{16}$$

where x, x', $\min(x)$, and $\max(x)$ indicate the original feature set, the normalized feature set, the minimum value of the feature set and the maximum value of the feature set, respectively. The min-max normalization in (16) does not affect the distribution of the features set. However, each feature of the above described feature sets is re-scaled to the same range from 0 to 1. Furthermore, we used these normalized unimodal feature sets to generate new multimodal feature sets, by using a feature-level fusion technique—a technique which serially concatenates feature vectors of the normalized feature sets that belong to the same activity sample.

B. CLASSIFICATION USING THE SVM

The SVM [55], [56] is an efficient supervised learning algorithm. The SVM is a binary classification algorithm that classifies the linearly separable data by determining an optimal hyperplane that best separates the labelled training examples (i.e., feature vectors) according to their class labels. In the terminology of the SVM, an optimal hyperplane is one that maximizes the margin between the closest training examples of both classes. The readers are referred to [55], [56] and the references therein for a succinct overview of SVM classification and regression. In situations, where the data are not linearly separable, the idea of "kernel induced feature space", also known as "the kernel trick", is used [56]. The kernel trick simply maps linearly inseparable data to a higher-dimensional space, where the data are more easily separable.

Furthermore, the SVM is only applicable to binary class problems. To overcome this limitation, there exist two approaches that enable binary classifiers to handle the multi-class problems, where the number of classes is more than two, namely one-versus-all (OvA) and one-versus-one (OvO).

In the OvA approach, we need to train a total of C number of binary classifiers for C number of classes (i.e., one classifier per class), where the data of the Cth class are labelled as the positive class and the data belonging to the rest of the classes are labelled as the negative class. In the testing phase, a test example is evaluated against all C classifiers and a confidence score is obtained per classifier. The class of the test example is determined by the corresponding classifier that has the highest confidence score.

In the OvO approach, first, an SVM classifier is trained for each pair of classes. This implies that for C number of classes, we must train $C(C - 1)/2$ classifiers. In the testing phase, a test example is evaluated against all trained classifiers, where each classifier predicts the class of the test example. The class that is predicted by a majority of classifiers is said to be the final predicted class of the test example.

C. TRAINING AND TESTING THE SVM

In supervised classification, first, labelled training data is used to train an SVM classifier/model. In the training phase, the classifier learns the parameters that map the feature vectors to their corresponding labels. Here, the aim is to train the model so well that later it can be used to predict the labels of new data that were not used in estimating the model. Therefore, after training a model, it is crucial to analyze its ability of the classifier to predict the labels of features vectors extracted from the unseen data. In this work, we used the K-fold cross-validation techniques [57] to estimate the performance of different models. In the K-fold cross-validation technique,

the feature vectors of a feature set (e.g., the CSI feature set) are randomly partitioned into K equal subsets. The features vectors of $K - 1$ subsets are used to train the model, and the features vectors of the remaining subset are used to test the performance of the trained model. The training process is repeated until each subset is used exactly once as validation data. The K results are then averaged to get an estimate about the model performance on future unseen data. Furthermore, we used the leave-one-subject-out (LOSO) cross-validation technique to access the participant-wise generalizability of the model. In the LOSO cross-validation technique, the feature vectors of one subject are used for testing the model, whereas the feature vectors of the remaining participants are used to train the model. We estimate the performance of the SVM classifier in terms of precision, recall, and accuracy metrics, which are defined as follows:

$$\text{Precision} = \frac{T_P}{T_P + F_P} \times 100\% \qquad (17)$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \times 100\% \qquad (18)$$

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \times 100\% \qquad (19)$$

where T_P, T_N, F_P, and F_N indicate true positive, true negative, false positive, and false negative, respectively.

True positive represents the number of the correctly classified examples of the positive class, whereas true negative represents the number of correctly classified examples of the negative class. False positive represents the number of examples that actually belong to the negative class, but they are misclassified as the positive class. False negative represents the number of examples that belong to the positive class in reality, but are misclassified as the negative class.

D. RESULTS AND DISCUSSION

Remember that there are four unimodal feature sets: the CSI-only, the accelerometer-only, the gyroscope-only, and the magnetometer-only feature set. From these four unimodal feature sets, we obtained six bimodal, four trimodal, and one quadmodal feature set. Within the scope of this work, the terms bi-, tri-, and quadmodal indicate that a feature set is obtained by combining two, three, and four different unimodal features sets, respectively. A complete list of different multimodal features sets that can be obtained by fusing the four unimodal features sets is presented in Table 2.

We used each unimodal and multimodal feature set to train and evaluate an SVM model. For simplicity and consistency, the names of the SVM models are derived from the names of the feature sets that are used to train and evaluate them. For instance, the SVM model trained and evaluated using the CSI-only feature set is called the unimodal CSI classifier. Similarly, the SVM model trained and evaluated using the accelerometer-only feature set is called the unimodal accelerometer classifier and so on. Each SVM classifier was trained using the OvO training strategy, and the 10-fold cross-validation technique was used to test the ability
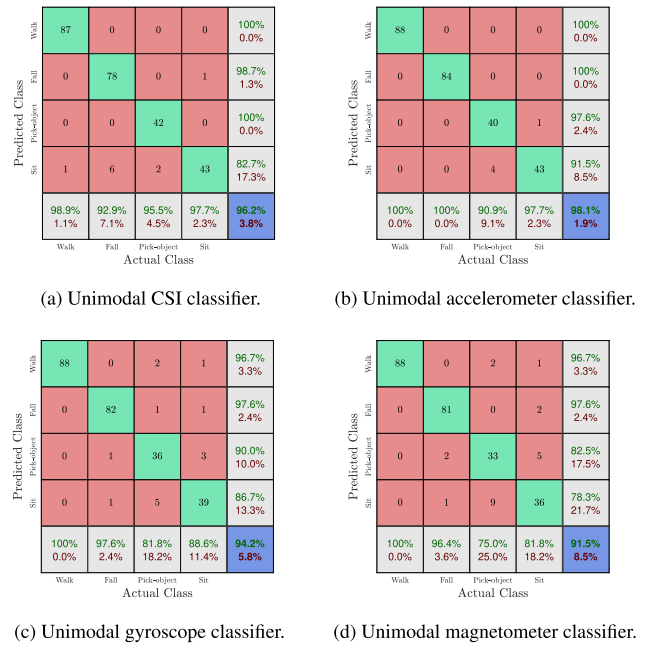

(a) Unimodal CSI classifier.   (b) Unimodal accelerometer classifier.

(c) Unimodal gyroscope classifier.   (d) Unimodal magnetometer classifier.

FIGURE 8. The confusion matrices of the unimodal classifiers.

TABLE 2. List of multimodal feature sets obtained by combining different unimodal feature sets. For simplicity, the names of the SVM models trained and evaluated using these feature sets are kept same as the names of the feature sets.

Feature set	CSI	Accelero-meter	Gyro-scope	Magneto-meter
Bimodal CA	×	×		
Bimodal CG	×		×	
Bimodal CM	×			×
Bimodal AG		×	×	
Bimodal AM		×		×
Bimodal GM			×	×
Trimodal CAG	×	×	×	
Trimodal CAM	×	×		×
Trimodal CGM	×		×	×
Trimodal AGM		×	×	×
Quadmodal CAGM	×	×	×	×

of the model to predict the unseen data. We employed the grid search algorithm to tune the hyper-parameters of each classifier. With the grid search algorithm, we found that each classifier reported the best validation results if the classifier is trained using the third-degree polynomial kernel.

We used the confusion matrices to describe the performance of each classifier. These confusion matrices are organized such that the rows and columns of each confusion matrix indicate the actual class and the predicted class, respectively. The diagonal green cells and the off-diagonal red cells present the number of correctly and incorrectly classified examples, respectively. The overall accuracy of the classifier is given in the blue cell of the confusion matrix. The far-right column and the last row of the confusion matrices present the precision and recall, respectively.
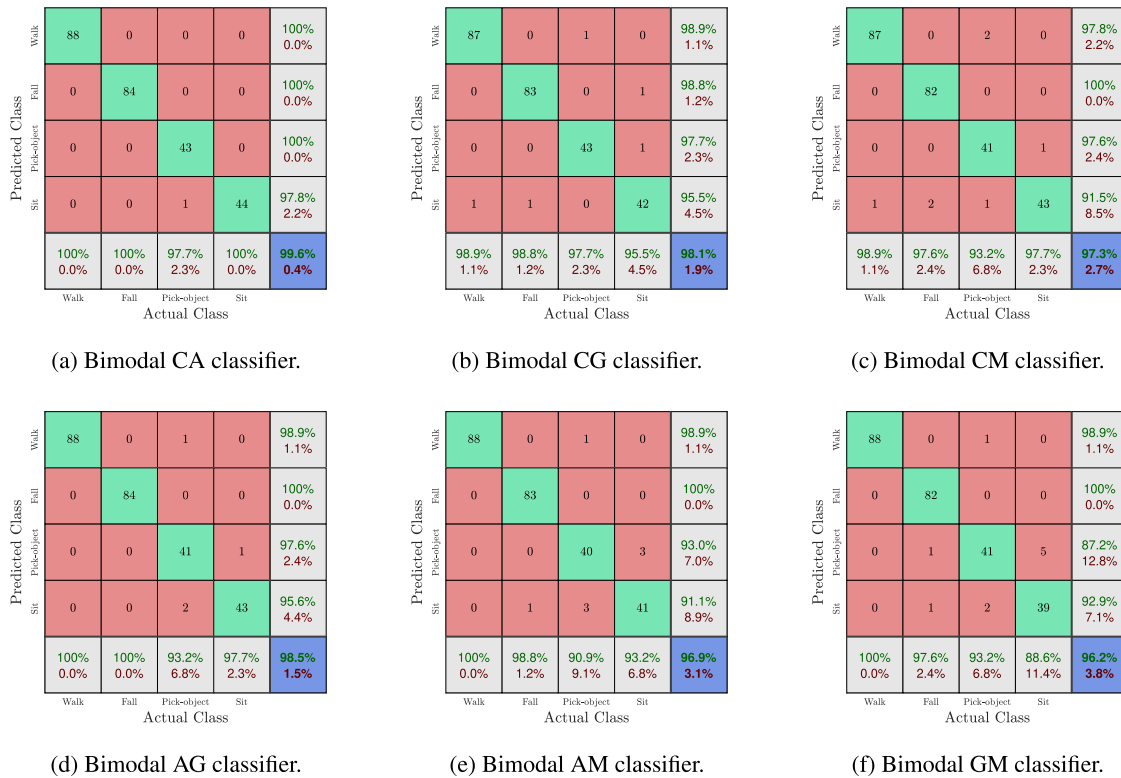
(a) Bimodal CA classifier.   (b) Bimodal CG classifier.   (c) Bimodal CM classifier.

(d) Bimodal AG classifier.   (e) Bimodal AM classifier.   (f) Bimodal GM classifier.

FIGURE 9. The confusion matrices of the bimodal classifiers.

1) RESULTS OF THE UNIMODAL CLASSIFIERS

The confusion matrices of the four unimodal SVM classifiers are presented in Fig. 8. The overall accuracy of the unimodal accelerometer classifier is 98.1% (see Fig. 8(b)), which is better than the overall recognition accuracy of the other three unimodal classifiers (see Figs. 8(a) 8(c), and 8(d)). The unimodal CSI classifier (see Fig. 8 (a)) achieves an overall accuracy of 96.2%, which is better than the unimodal gyroscope classifier (see Fig. 8 (c)) and the unimodal magnetometer classifier (see Fig. 8(d)). The unimodal accelerometer classifier recognizes the walking and falling activities with 100% precision and recall. However, it mixes up the sitting and picking up an object activities. The unimodal CSI classifier recognizes the walking activity with 100% precision and 98.8% recall. For the falling, picking up an object, and sitting activities, the unimodal CSI classifier achieves 98.7%, 100%, and 82.7% precision score, respectively. The recall or the sensitivity score of falling, picking up an object, and sitting activities is 92.9%, 95.5%, and 97.7%, respectively. Moreover, we can observe that the unimodal gyroscope and magnetometer classifiers mixes up all four activities, and they have an overall recognition accuracy of 94.2% and 91.5%, respectively.

2) RESULTS OF THE BIMODAL CLASSIFIERS

The results of the bimodal classifiers are presented in Fig. 9. The bimodal classifiers are trained and evaluated using the bimodal feature sets obtained by combining two different unimodal feature sets.

We can observe that the bimodal CA classifier outperforms all other bimodal classifiers presented in Figs. 9(b)–9(f) with an overall accuracy of 99.6%. This indicates that the bimodal CA classifier performs better the unimodal CSI-only and the unimodal accelerometer-only classifiers. The bimodal CA classifier only makes a single prediction error by wrongly predicting the label of an activity sample of the picking up an object activity as the sitting activity. Thus, the recall score of the picking activity is 97.7%, whereas the recall score of the other three activities is 100%.

Moreover, the bimodal CA classifier recognizes the three activities: walking, falling, and picking up an object with 100% precision and the sitting activity with 97.8% precision. The performance of the bimodal AG classifier (see Fig. 9(d)) is better than the unimodal accelerometer-only and the unimodal gyroscope-only classifiers. Similarly, the bimodal CG, CM, and GM classifiers follow the same trend.

The results of the bimodal classifiers provide the following three significant insights: (1) the fusion of the CSI-based feature set with any of the accelerometer-, gyroscope- or the magnetometer-based feature set significantly improves the overall recognition accuracy compared to the corresponding individual feature-set-based classification results, (2) the fusion of the accelerometer- and gyroscope-based feature set slightly improves the recognition accuracy, similarly the
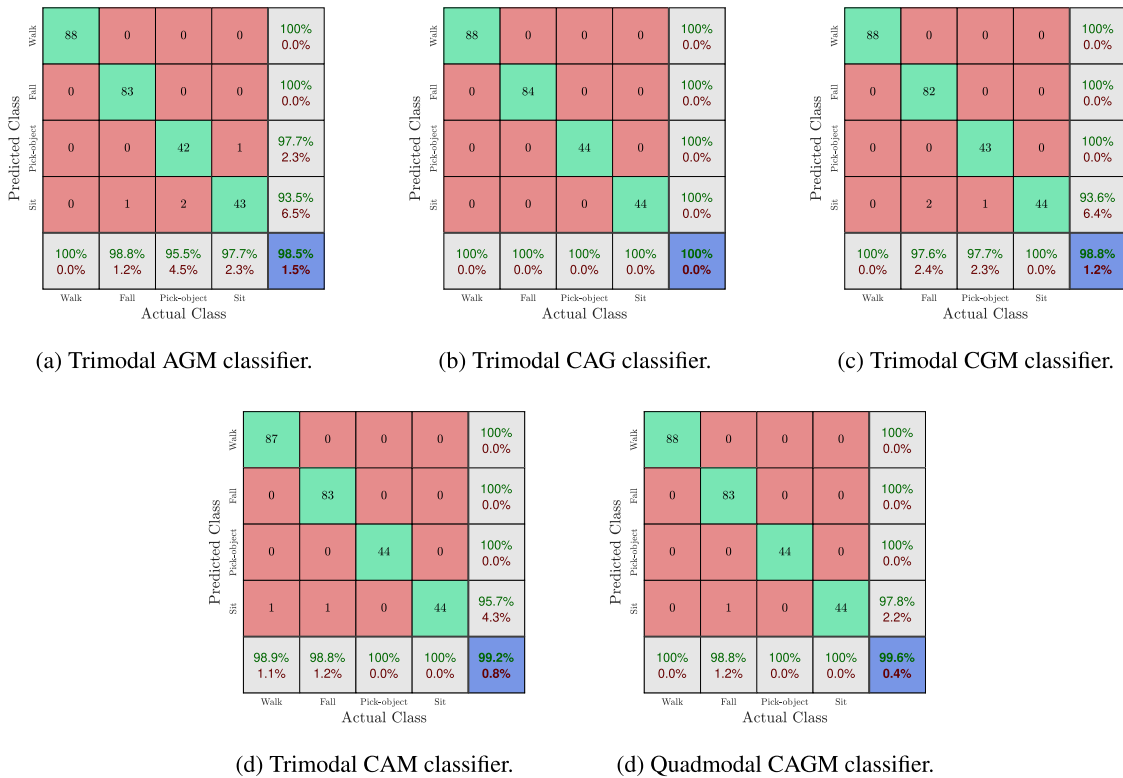
(a) Trimodal AGM classifier.

(b) Trimodal CAG classifier.

(c) Trimodal CGM classifier.

(d) Trimodal CAM classifier.

(d) Quadmodal CAGM classifier.

FIGURE 10. The confusion matrices of the trimodal and quadmodal classifiers.

fusion of the gyroscope- and magnetometer-based feature sets also improves the overall recognition accuracy; and (3) as reported in the previous work [58], our results of the bimodal AM classifier also indicate that the fusion of the accelerometer and the magnetometer-based features sets do not improve recognition accuracy.

3) RESULTS OF THE TRIMODAL AND QUADMODAL CLASSIFIERS

The results of the trimodal classifiers are presented in Fig. 10. The trimodal classifiers are trained and evaluated using the trimodal feature sets obtained by combining three different unimodal feature sets.

We can observe in Fig. 10(b) that the trimodal CAG classifier outperforms the other three trimodal classifiers presented in Fig. 10. The trimodal CAG classifier recognizes each activity with 100% precision and recall, thus its overall recognition accuracy is 100%. The trimodal CAM classifier (see Fig. 10(d)) that uses the features set obtained by combining the unimodal CSI-only, accelerometer-only, and magnetometer-only feature sets, achieves an overall recognition accuracy of the 99.2%, which is slightly lower than the bimodal CA classifier (see Fig. 9(a)). The results of the trimodal classifiers further confirms the conclusion drawn based on the results of the bimodal classifiers. For instance, the trimodal AGM classifier achieves an overall recognition accuracy of 98.5%, which is similar to the overall recognition

accuracy of the bimodal AG classifier. Similarly, the performance of the trimodal CAM classifier is not better than the bimodal CA classifier. Moreover, we also notice the same trend as the performance of the quadmodal CAGM classifier is slightly lower than the performance of the trimodal CAG classifier. This advocates that combining the magnetometer-based feature set either with the stand-alone accelerometer-based feature set or with a feature set that already contains an accelerometer-based feature set does not improve the classification accuracy.

E. SUBJECT-WISE GENERALIZABILITY

We used the LOSO cross-validation technique to assess the subject-wise generalizability of the proposed multimodal WiWeHAR. For this purpose, we used the best performing feature set, i.e., the trimodal CAG feature set which is obtained by fusing the CSI-, accelerometer-, and gyroscope-based feature sets. The feature vectors of the trimodal CAG feature set that correspond to one subject are used for testing the SVM classifier, which was trained by using the remaining feature vectors of the trimodal CAG feature set. We repeated this process until the feature vectors of each subject are exactly once used as the testing data. The results of the LOSO cross-validation technique are presented in Fig. 11. We can observe that the activities performed by five out of the nine subjects can be recognized with 100% accuracy, whereas the overall recognition accuracy for subjects 1, 2, 3,
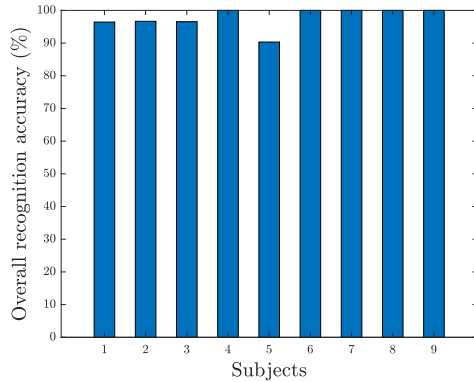
FIGURE 11. Subject-wise recognition accuracy of the best performing trimodal CAG classifier.

and 5 is 96.43%, 96.67%, 96.55%, and 90.32%, respectively. A possible reason of low recognition accuracy of Subject 5 is that each participant has her/his own way of performing the same activity. Therefore, it is possible that a classifier trained using the data associated with other persons might not incorporate some user specific variations and thus results in low recognition results. This issue can easily be solved either by collecting data from more subjects or by retraining the classifiers for subjects with low recognition results by using their data in the training phase.

VII. CONCLUSION AND FUTURE WORK

In this article, we proposed WiWeHAR– an activity recognition system that consists of multimodal sensing and machine learning phases. In the sensing phase, WiWeHAR uses standard Wi-Fi T_x and R_x devices, and an IMU to simultaneously collect the Wi-Fi CSI and the inertial data, respectively. The IMU consists of an accelerometer, a gyroscope, and a magnetometer sensor. The multimodal data were collected from nine participants while performing four different activities. These activities involved walking, falling on a mattress, picking up an object from the floor, and sitting on a chair. The collected CSI and inertial data were processed to reduce the impact of noise. Thereafter, we extracted time- and frequency-domain features individually from the MDS, accelerometer magnitude, gyroscope magnitude, magnetometer magnitude data. Moreover, we generated bi-, tri-, and quadmodal features set by fusing the unimodal feature sets extracted from the MDS accelerometer magnitude, gyroscope magnitude, and magnetometer magnitude data. Thereafter, we trained and evaluated multiple SVM classifiers by using the uni-, bi-, tri-, and quadmodal feature sets. Our results show that the multimodal approach achieved an overall accuracy of 99.6%–100%, whereas the unimodal CSI, accelerometer, gyroscope, and magnetometer approach achieved overall accuracies of 96.2%, 98.1%, 94.2%, and 91.5%, respectively. Furthermore, the multimodal approach also outperforms the other four unimodal approaches in terms of precision and recall. Based on these experimental results, we can conclude that a highly accurate and robust HAR system can

be designed by incorporating Wi-Fi and wearable sensing modalities. In the future, we will extend multimodal fusion by incorporating more sensing modalities, such as acoustic sensors. We will use a massive MIMO system combined with multimodal approach to recognize complex human activities in more complex scenarios. Moreover, we will also study the influence of antenna quality on the the performance of CSI system.

APPENDIX

In this section, we present a brief overview of the 23 time- and frequency-domain features that are extracted from the active MDS and active magnitude segments. The first four features are simple statistical measures, such as mean, variance, standard deviation, and skewness whereas, a brief description and mathematical definitions of the remaining 19 features are given below.

1) MEAN ABSOLUTE VALUE (MAV)

MAV computes the average absolute value of the active segment as

$$MAV = \frac{1}{N} \sum_{n=1}^{N} |x_n|. \tag{20}$$

2) WAVEFORM LENGTH [59], [60]

The waveform length is the cumulative length of the active segment waveform and is expressed as

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n|. \tag{21}$$

3) ENHANCED MEAN ABSOLUTE VALUE (EMAV) [61]

It is an extended version of the MAV and defined by

$$EMAV = \frac{1}{N} \sum_{n=1}^{N} |(x_n)^\alpha|,$$

$$\text{where} \quad \alpha = \begin{cases} 0.75, & \text{if } 0.2N \leq n \leq 0.8N \\ 0.5, & \text{otherwise.} \end{cases} \tag{22}$$

4) ENHANCED WAVEFORM LENGTH [61]

It is an extended version of the waveform length defined by

$$EWL = \sum_{n=1}^{N-1} |(x_{n+1} - x_n)^\alpha|,$$

$$\text{where} \quad \alpha = \begin{cases} 0.75, & \text{if } 0.2N \leq n \leq 0.8N \\ 0.5, & \text{otherwise.} \end{cases} \tag{23}$$

5) WEIGHTED MEAN ABSOLUTE VALUE I (WMAVI) [59], [60]

The WMAVI is an extension of the mean absolute value, where a weighted window function is introduced to the equation of the MAV according to

$$MMAVI = \frac{1}{N} \sum_{n=1}^{N} w_n |x_n|,$$

$$\text{where} \quad w_n = \begin{cases} 1.0, & \text{if } 0.25N \leq n \leq 0.75N \\ 0.5, & \text{otherwise.} \end{cases} \quad (24)$$

6) WEIGHTED MEAN ABSOLUTE VALUE II (WMAVII) [59], [60]

Similar to the WMAVI, the WMAVII is another extension of the MAV. In contrast to WMAVI, the window function in WMAVII is continuous. Mathematically, the WMAVII is defined as

$$MMAVII = \frac{1}{N}\sum_{n=1}^{N} w_n|x_n|,$$

$$\text{where} \quad w_n = \begin{cases} 1.0, & \text{if } 0.25N \leq n \leq 0.75N \\ \frac{4n}{N}, & \text{if } n < 0.25N \\ \frac{4(n-N)}{N}, & \text{otherwise.} \end{cases} \quad (25)$$

7) MAXIMUM FRACTAL LENGTH (MFL) [60]

The MFL measures the small changes in the complexity of the active segment.

$$MFL = \log\sqrt{\sum_{n=1}^{N-1}(x_{n+1} - x_n)^2}. \quad (26)$$

8) MEAN AMPLITUDE CHANGE (MAC) [60]

The MAC indicates the mean cumulative length of the waveform of the active segment, which is given by

$$AAC = \frac{1}{N}\sum_{n=1}^{N-1}|x_{n+1} - x_n|. \quad (27)$$

9) ROOT MEAN SQUARE (RMS)

The RMS feature indicates the average magnitude of the active segment. Mathematically, the RMS is expressed as

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N}x_n^2}. \quad (28)$$

10) DIFFERENCE ABSOLUTE STANDARD DEVIATION (DASDV) [60]

The DASDV computes the standard deviation of the wavelength according to

$$DASDV = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N-1}(x_{n+1} - x_n)^2}. \quad (29)$$

11) SIMPLE SQUARED INTEGRAL [61]

It indicates the energy index of the active segment. Mathematically, this feature is given by

$$SSI = \sum_{i=1}^{N}x_n^2. \quad (30)$$

12) WILLISON AMPLITUDE (WAMP) [60], [61]

The WAMP counts the number of times the absolute difference between two consecutive values of the active segment is higher than a certain threshold value. This feature is defined as

$$WAMP = \sum_{n=1}^{N}f(|x_{n+1} - x_n|),$$

$$\text{where} \quad f(|x_{n+1} - x_n|) = \begin{cases} 1, & \text{if } x > \text{threshold} \\ 0, & \text{otherwise.} \end{cases} \quad (31)$$

13) ZERO CROSSING [59], [60]

This feature shows how often the waveform of the active segment has changed its sign. The mathematical definition of zero crossing is given by

$$\{x_{n+1} < 0 \wedge x_n > 0\} \vee \{x_{n+1} > 0 \wedge x_n < 0\}$$
$$\wedge |x_{n+1} - x_n| \geq \text{threshold.} \quad (32)$$

14) SLOPE SIGN CHANGE (SSC) [59], [60]

This feature counts the number of times the sign of the slope of the waveform of the active segment is changed. Given the three consecutive samples of an active segment, the change in sign of the slope occurs if

$$[\{x_n > x_{n+1} \wedge x_n > x_{n-1}\} \vee \{x_n < x_{n+1} \wedge x_n < x_{n-1}\}] \wedge$$
$$|x_n - x_{n+1}| \geq \text{threshold} \vee |x_n - x_{n-1}| \geq \text{threshold.}$$

15) MAX OF THE ABSOLUTE VALUE

For a given active segment, we compute its absolute value $|x|$, and then determine the largest value of $|x|$.

16) SLOPE

For a given active segment, we first determine the index i_1 at which x reaches its maximum absolute value $|x|$ max. Then we select an interval $[i_{f0}, i_{f1}]$ located right before i_1 and determine the slope of x within the interval $[i_{f0}, i_{f1}]$. To this end, we fit the values of x for $i \in [i_{f0}, i_{f1}]$ with a linear function and then compute the slope of this line. For a fall activity the slope is large compared to the slope of walking.

17) PEAKS OF THE AUTOCORRELATION

For a given active segment, we compute its autocorrelation function (ACF) $R_x(m)$, defined as

$$R_x(m) = \frac{1}{2N+1}\sum_{n=-N}^{N}x_n^* x_{n+m} \quad (33)$$

where N is the length of the active segment. From the ACF $R_x(m)$, we determine the locations and values of the first and the second peaks.

18) SPECTRAL PEAKS

The power spectral density (PSD) $S_x(f)$ of an active segment can be computing by taking the Fourier transform of the ACF

$R_x(m)$ as

$$S_x(f) = F\{R_x(m)\} = \sum_{m=-\infty}^{\infty} R_x(m)e^{-j2\pi mfT_S} d\tau, \quad (34)$$

where $1/T_S$ is the sampling rate of the active segment. From the PSD $S_x(f)$, we extract the values and locations associated with the 10 main peaks. This feature enables identifying the main harmonics embedded in x, which captures a quantitative information on the time variation of x.

19) SPECTRAL ENERGY

This feature captures the energy in different frequency bands of the active segment. To extract this feature, we first compute the PSD $S_x(f)$. Then we divide the spectrum of $S_x(f)$ into ten frequency bands and evaluate the energy contained in each frequency band.

REFERENCES

[1] S. Ranasinghe, F. A. Machot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 8, pp. 1–22, 2016.

[2] P. Menschner, A. Prinz, P. Koene, F. Köbler, M. Altmann, H. Krcmar, and J. M. Leimeister, "Reaching into patients' homes–participatory designed AAL services," *Electron. Markets*, vol. 21, no. 1, pp. 63–76, Feb. 2011.

[3] D. Podsiadlo and S. Richardson, "The timed 'up & go': A test of basic functional mobility for frail elderly persons," *J. Amer. Geriatrics Soc.*, vol. 39, no. 2, pp. 142–148, 1991.

[4] L. Z. Rubenstein, K. R. Josephson, G. D. Wieland, P. A. English, J. A. Sayre, and R. L. Kane, "Effectiveness of a geriatric evaluation unit: A randomized clinical trial," *New England J. Med.*, vol. 311, no. 26, pp. 1664–1670, 1984.

[5] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *J. Amer. Geriatrics Soc.*, vol. 66, no. 4, pp. 693–698, Apr. 2018.

[6] E. Chiauzzi, C. Rodarte, and P. DasMahapatra, "Patient-centered activity monitoring in the self-management of chronic health conditions," *BMC Med.*, vol. 13, no. 1, pp. 1–6, Dec. 2015.

[7] J. Clemente, F. Li, M. Valero, and W. Song, "Smart seismic sensing for indoor fall detection, location, and notification," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 524–532, Feb. 2020.

[8] F. Fusier, V. Valentin, F. Brémond, M. Thonnat, M. Borg, D. Thirde, and J. Ferryman, "Video understanding for complex activity recognition," *Mach. Vis. Appl.*, vol. 18, nos. 3–4, pp. 167–188, May 2007.

[9] P. Peursum, G. West, and S. Venkatesh, "Combining image regions and human activity for indirect object recognition in indoor wide-angle views," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 82–89.

[10] M.-C. Chang, N. Krahnstoever, S. Lim, and T. Yu, "Group level activity recognition in crowded environments across multiple cameras," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2010, pp. 56–63.

[11] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations," in *Proc. 7th Int. Conf. Ubiquitous Intell. Comput.*, 2010, pp. 548–562.

[12] J. Guo, X. Zhou, Y. Sun, G. Ping, G. Zhao, and Z. Li, "Smartphone-based patients' activity recognition by using a self-learning scheme for medical monitoring," *J. Med. Syst.*, vol. 40, no. 6, p. 140, Jun. 2016.

[13] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Gener. Comput. Syst.*, vol. 81, pp. 307–313, Apr. 2018.

[14] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, "A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities," in *Proc. IEEE 12th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Jun. 2015, pp. 1–6.

[15] G. Uslu, H. I. Dursunoglu, O. Altun, and S. Baydere, "Human activity monitoring with wearable sensors and hybrid classifiers," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 5, pp. 345–353, 2013. [Online]. Available: http://www.mirlabs.org/ijcisim/index.php

[16] M. Georgi, C. Amma, and T. Schultz, "Recognizing hand and finger gestures with IMU based motion and EMG based muscle activity sensing," in *Proc. Int. Conf. Bio-Inspired Syst. Signal Process.*, 2015, pp. 99–108.

[17] F. Lorussi, N. Carbonaro, D. De Rossi, R. Paradiso, P. Veltink, and A. Tognetti, "Wearable textile platform for assessing stroke patient treatment in daily life conditions," *Frontiers Bioeng. Biotechnol.*, vol. 4, p. 28, Mar. 2016.

[18] Carre Technologies Inc (Hexoskin). *Hexoskin Smart Garments Specifications*. Accessed: May 5, 2020. [Online]. Available: https://www.hexoskin.com/collections/all

[19] K. Kunze and P. Lukowicz, "Sensor placement variations in wearable activity recognition," *IEEE Pervas. Comput.*, vol. 13, no. 4, pp. 32–41, Oct. 2014.

[20] N. Kale, J. Lee, R. Lotfian, and R. Jafari, "Impact of sensor misplacement on dynamic time warping based human activity recognition using wearable computers," in *Proc. Conf. Wireless Health (WH)*, 2012, pp. 1–8.

[21] O. D. Incel, "Analysis of movement, orientation and rotation-based sensing for phone placement recognition," *Sensors*, vol. 15, no. 10, pp. 25474–25506, 2015.

[22] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A review on human activity recognition using vision-based method," *J. Healthcare Eng.*, vol. 2017, pp. 1–31, Jul. 2017.

[23] 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.

[24] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Apr. 2014.

[25] J. Chen, F. Li, H. Chen, S. Yang, and Y. Wang, "Dynamic gesture recognition using wireless signals with less disturbance," *Pers. Ubiquitous Comput.*, vol. 23, no. 1, pp. 17–27, Feb. 2019.

[26] D. Palipana, "Fall detection using channel state information from WiFi devices," Ph.D. dissertation, Cork Inst. Technol., Cork, Ireland, 2019. [Online]. Available: https://sword.cit.ie/engdiss/6

[27] T. Z. Chowdhury, "Using Wi-Fi channel state information (CSI) for human activity recognition and fall detection," M.S. thesis, Dept. Elect. Comput. Eng., Univ. British Columbia, Vancouver, BC Canada, 2018.

[28] H. Li, X. He, X. Chen, Y. Fang, and Q. Fang, "Wi-motion: A robust human activity recognition using WiFi signals," *IEEE Access*, vol. 7, pp. 153287–153299, 2019.

[29] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.

[30] 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, May 2019.

[31] M. Scholz, T. Riedel, M. Hock, and M. Beigl, "Device-free and device-bound activity recognition using radio signal strength," in *Proc. 4th Augmented Hum. Int. Conf. (AH)*, 2013, pp. 100–107.

[32] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2014, pp. 617–628.

[33] J. Ding and Y. Wang, "WiFi CSI-Based human activity recognition using deep recurrent neural network," *IEEE Access*, vol. 7, pp. 174257–174269, 2019.

[34] W. Elmenreich, "Sensor fusion in time-triggered systems," Ph.D. dissertation, Fakultät für Technische Naturwissenschaften und Informatik, Technische Universität Wien, Wien, Austria, 2002.

[35] H. Zou, J. Yang, H. P. Das, H. Liu, Y. Zhou, and C. J. Spanos, "WiFi and vision multimodal learning for accurate and robust device-free human activity recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 326–433.

[36] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong, "Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning," *Sensors*, vol. 19, no. 7, p. 1716, Apr. 2019.

[37] Y. Shu, C. Chen, K.-I. Shu, and H. Zhang, "Research on human motion recognition based on Wi-Fi and inertial sensor signal fusion," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 496–504.

[38] R. Ramezani, Y. Xiao, and A. Naeim, "Sensing-Fi: Wi-Fi CSI and accelerometer fusion system for fall detection," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 402–405.

[39] Clinical Trial Transformation Initiative. (2017). *CTTI Recommendations: Developing Novel Endpoints Generated by Mobile Technology For Use in Clinical Trials*. Accessed: May 5, 2020. [Online]. Available: https://www.ctti-clinicaltrials.org/files/novelendpoints-recs.pdf

[40] Intel. *Product Brief Intel Ultimate N WiFi Link 5300*. Accessed: May 5, 2020. [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/ultimate-n-wifi-link-5300-brief.pdf

[41] A. Abdelgawwad and M. Paetzold, "A framework for activity monitoring and fall detection based on the characteristics of indoor channels," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–7.

[42] A. Abdelgawwad and M. Patzold, "On the influence of walking people on the Doppler spectral characteristics of indoor channels," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–7.

[43] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.

[44] 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.

[45] H. Zhu, Y. Zhuo, Q. Liu, and S. Chang, "π-splicer: Perceiving accurate CSI phases with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 17, no. 9, pp. 2155–2165, Sep. 2018.

[46] N. Keerativoranan, A. Haniz, K. Saito, and J.-I. Takada, "Mitigation of CSI temporal phase rotation with B2B calibration method for fine-grained motion detection analysis on commodity Wi-Fi devices," *Sensors*, vol. 18, no. 11, p. 3795, Nov. 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/11/3795

[47] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," in *Proc. 20th IEEE Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2014, pp. 1–8.

[48] A. Abdelgawwad, A. Borhani, and M. Pätzold, "Modelling, analysis, and simulation of the micro-Doppler effect in wideband indoor channels with confirmation through pendulum experiments," *Sensors*, vol. 20, no. 4, p. 1049, 2020.

[49] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, Sep. 2019.

[50] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 2002.

[51] A. D. Papaioannou, "Component analysis of complex-valued data for machine learning and computer vision tasks," Ph.D. dissertation, Imperial College London, London, U.K., 2017.

[52] B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2015.

[53] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: From authentication to imitation," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3209–3221, Nov. 2017.

[54] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.

[55] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[56] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[57] F. Mosteller and J. W. Tukey, "Data analysis, including statistics," in *Handbook of Social Psychology*, vol. 2. Reading, MA, USA: Addison-Wesley, 1968.

[58] C. V. San Buenaventura and N. M. C. Tiglao, "Basic human activity recognition based on sensor fusion in smartphones," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manage. (IM)*, May 2017, pp. 1182–1185.

[59] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, Aug. 2008.

[60] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[61] H. A. Elkholy, A. T. Azar, A. Magd, H. Marzouk, and H. H. Ammar, "Classifying upper limb activities using deep neural networks," in *Proc. Int. Conf. Artif. Intell. Comput. Vis. (AICV)*, 2020, pp. 268–282.

MUHAMMAD MUAAZ received the M.Sc. degree in information and communication systems security from the KTH Royal Institute of Technology, Sweden, and the Ph.D. degree in computer science from Johannes Keppler University, Austria. He is currently a Postdoctoral Fellow with the University of Agder, Norway, where he is working towards unobtrusive human activity recognition methods using Wi-Fi and wearable sensors. His research interests include information security, data analysis, machine learning, deep learning, and biometrics.

ALI CHELLI (Member, IEEE) received the B.Sc. degree in communications from the Ecole Superieure des Communications de Tunis (SUP'COM), in 2005, and the M.Sc. and Ph.D. degrees in information and communication technology from the University of Agder, Norway, in 2007 and 2013, respectively. He was serving as a Postdoctoral Fellow with the King Abdullah University of Science and Technology (KAUST) and the Norwegian Institute of Science and Technology (NTNU), Trondheim, Norway. He is currently a Postdoctoral Fellow with the University of Agder. His research interests include machine learning, deep learning, wireless communication theory, cooperative relaying, vehicle-to-vehicle communications, game theory, and channel modeling. He has been serving as a TPC member and a TPC Chair in several conferences.

AHMED ABDELMONEM ABDELGAWWAD was born in Minya, Egypt, in 1991. He received the bachelor's and master's degrees (Hons.) in information engineering and technology from German University in Cairo (GUC). His master's degree was related to the complexity reduction of 3D ray-tracing for indoor coverage solutions. He is currently pursuing the Ph.D. degree with the University of Agder. His research interests and background include network modeling and simulation, channel modeling for fall detection systems, non-stationary channel models, and time-frequency analysis for non-stationary channel models.

ANDREU CATALÀ MALLOFRÉ received the B.Sc./M.Sc. degrees in physics from the University of Barcelona and the Ph.D. degree in physics from the Technical University of Catalonia. From 2000 to 2006, he worked as the Dean of the Engineering School, Vilanova i la Geltrú. He is currently a Full Professor with the Department of Automatic Control, Technical University of Catalonia, and the Director of the Research Centre for Dependency Care and Autonomous Living. His research interests include assistive technologies, artificial intelligence, social robotics, and wearable and intelligent sensors. He is a Coordinator of international and national research projects, among them the European Thematic Network in Falls Prevention Intervention and Security (E-NOFALLS 2013–2016). He is a member of the coordination team of the EIPAHA (Action Group A2 in Personalised HEALTH). He is a promoter of a spin-off called SENSE4CARE.

MATTHIAS PÄTZOLD (Senior Member, IEEE) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Ruhr University Bochum, Bochum, Germany, in 1985 and 1989, respectively, and the Habilitation degree in communications engineering from the Hamburg University of Technology, Hamburg, Germany, in 1998. From 1990 to 1992, he worked for ANT Nachrichtentechnik GmbH, Backnang, Germany, where he was involved in digital satellite communications. From 1992 to 2001, he was with the Department of Digital Networks, Hamburg University of Technology. Since 2001, he has been a Full Professor of mobile communications with the University of Agder, Norway. He has authored several books and numerous technical articles. He has been actively participating in numerous conferences as a TPC member and a TPC Chair. His publications received 14 best paper awards.

● ● ●