



Wi-Sense: a passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems

Muhammad Muaaz¹ · Ali Chelli¹ · Martin Wulf Gerdes¹ · Matthias Pätzold¹

Received: 29 August 2020 / Accepted: 18 June 2021 / Published online: 13 July 2021
© The Author(s) 2021

Abstract

A human activity recognition (HAR) system acts as the backbone of many human-centric applications, such as active assisted living and in-home monitoring for elderly and physically impaired people. Although existing Wi-Fi-based human activity recognition methods report good results, their performance is affected by the changes in the ambient environment. In this work, we present Wi-Sense—a human activity recognition system that uses a convolutional neural network (CNN) to recognize human activities based on the environment-independent fingerprints extracted from the Wi-Fi channel state information (CSI). First, Wi-Sense captures the CSI by using a standard Wi-Fi network interface card. Wi-Sense applies the CSI ratio method to reduce the noise and the impact of the phase offset. In addition, it applies the principal component analysis to remove redundant information. This step not only reduces the data dimension but also removes the environmental impact. Thereafter, we compute the processed data spectrogram which reveals environment-independent time-variant micro-Doppler fingerprints of the performed activity. We use these spectrogram images to train a CNN. We evaluate our approach by using a human activity data set collected from nine volunteers in an indoor environment. Our results show that Wi-Sense can recognize these activities with an overall accuracy of 97.78%. To stress on the applicability of the proposed Wi-Sense system, we provide an overview of the standards involved in the health information systems and systematically describe how Wi-Sense HAR system can be integrated into the eHealth infrastructure.

Keywords Channel state information · Convolutional neural network · Doppler effect · Health information systems · Human activity recognition · Principal component analysis · Radio frequency sensing · Spectrogram

1 Introduction

The world demographics reveal that the elderly population is rapidly increasing across the globe. The World Health Organization statistics show that 16% of the world population will

be over the age of 65 by 2050 [34]. To embrace this demographic shift and to poise the social transformation associated with it, the Madrid International Plan of Action on Ageing [33] has identified several priority directions including “ensuring enabling and supportive environments.” Issue I of this priority direction emphasizes the development of adaptable living environments for elderly to support them to enjoy safe, active, and independent living in their homes for as long as possible. To enable this, there is a great need to develop robust, unobtrusive, and senior-friendly in-home monitoring systems that can be integrated with HIS to automatically invoke a nearby emergency health service provider for assistance in case of emergency.

HAR acts as the basic building block of an in-home monitoring system. HAR generally deals with the recognition of human activities based on sensor data. A typical HAR system consists of data sensing, processing, and classification modules. The sensing module consists of one or more sensors that collect data when a person is carrying

✉ Muhammad Muaaz
muhammad.muaaz@uia.no

Ali Chelli
ali.chelli@uia.no

Martin Wulf Gerdes
martin.gerdes@uia.no

Matthias Pätzold
matthias.paetzold@uia.no

¹ Faculty of Engineering and Science, University of Agder, P.O. Box 509, 4898 Grimstad, Norway

out different activities. A data processing module generally eliminates noise from the raw sensor data and subsequently prepares the data for the classification module. The classification module classifies the performed activities using a learning algorithm.

HAR systems are generally classified into vision-based [3, 19], wearable sensor-based [32, 38], and radio frequency (RF)-based [12, 37]. In vision-based HAR systems, computer vision techniques are used to recognize human activities from the recorded videos or images [43]. Vision-based HAR systems are generally considered very accurate in recognizing human activities, but they do not cope well with changes in the ambient environment. For example, vision-based systems have a limited operating area; they require a clear (or an obstacle-free) view of the environment; users often rate them as a potential privacy risk; their performance is subject to change under different lighting conditions and anthropometric variations. On the other hand, wearable sensor-based HAR systems use inertial sensors to capture the user's dynamic body movements while performing different activities. The recorded sensor data is then processed and analyzed to recognize activities. Wearable sensor-based HAR systems can be considered as a viable alternative to vision-based HAR systems because they are immune to changes in the ambient environment and do not suffer from privacy risks. However, wearable sensor-based HAR systems are less user-friendly because they require users to carry or wear sensors, which is invasive, unpleasant, and uncomfortable for the elderly and physically disabled individuals [23].

RF-based HAR systems work on the principle that human bodies reflect RF signals, and human movements introduce variations in the frequencies of the RF signals due to the physical phenomenon known as the Doppler effect. These variation-enriched RF signals are used to recognize human activities. RF-based HAR systems offer several advantages over vision-based and wearable sensor-based HAR techniques. For instance, in contrast to vision-based HAR systems, RF-based HAR systems are unaffected by lighting conditions as well as anthropometric variations, and they do not compromise the user's privacy. In addition to that, unlike wearable sensor-based HAR systems, RF-based HAR systems do not restrict users to wear a sensor. Therefore, RF-based HAR systems can be considered as a more appropriate alternative for HAR systems in smart homes and healthcare applications. As a result, since last few years, researchers have been diligently exploring and developing RF-based HAR techniques.

1.1 Related work

Existing RF-based HAR systems either employ radar or Wi-Fi technology to sense human activities, both of which

have their own merits and demerits. For instance, radar-based systems [22] offer high sensitivity and spatial resolution resulting in a more accurate recognition of human activities [36], gestures [31], and fall detection [16], but this advantage comes at a significantly higher cost. Therefore, the high cost of radar-based systems limits their widespread use. On the other hand, Wi-Fi-based HAR systems are economical in general, and they can easily be integrated into pre-existing Wi-Fi infrastructures at our homes and workplaces without notable additional costs. There exist two types of Wi-Fi-based HAR systems [37]. The first type uses the received signal strength indicator (RSSI) [12, 29], whereas the second type relies on the channel state information (CSI) [7, 8, 15, 25, 37, 40, 44] for activity recognition tasks.

The CSI characterizes how RF signals travel from a transmitter to a receiver in an environment at different carrier frequencies and undergo various effects, such as amplitude attenuation, phase shift, and time delay [2]. The previous work has shown that the CSI-based HAR systems outperform the RSSI-based HAR systems [37]. This is due to a reason that the CSI provides more information compared to the RSSI. For each received data packet, the CSI provides both the amplitude and phase information for every orthogonal frequency-division multiplexing (OFDM) subcarrier, whereas the RSSI provides a single-value for each received packet that represents the attenuation of the received signal strength during propagation.

There exist various approaches to recognize human activities using CSI data by applying machine learning [21, 25, 37] and deep learning [7, 44] techniques. In [37], the authors proposed two theoretical models. The first model (known as “the CSI-speed model”) links the speed of human body movements with the CSI data, while the second model (known as “the CSI-activity model”) links the speed of human body movements with human activities [37]. The proposed approach in [37] was developed using commercial Wi-Fi devices and achieved an overall recognition accuracy of 96%. A deep learning technique consisting of autoencoder, convolutional neuronal network (CNN), and long short-term memory modules to recognize human activities from the CSI data has been proposed in [44]. This deep learning network achieved an overall accuracy of 97.4%. In [7], an attention-based bidirectional long short-term memory technique was used to recognize human activities from the CSI data. The CSI data sets collected in two different environments, namely an activity room and a meeting room were used to evaluate the performance of the proposed approach. This approach achieved a recognition accuracy of 96.7% when evaluated using the CSI data collected inside the activity room and 97.3% when the CSI data collected inside the meeting room was used. However, in the cross-environment scenario, where the training data that has been

collected in one environment and the testing data from the other environment are used, the overall recognition accuracy drops to 32%.

Although existing CSI-based HAR systems have reported reasonably good results, they still suffer from the drawback that they are environment dependent. This implies that their performance is susceptible to changes in the environment, as reported in [7]. This is generally true for all CSI based-HAR systems that only rely on the amplitude of the CSI data for recognizing activities and gestures. Researchers have proposed several solutions to resolve this problem. For instance, in [15], it was proposed to use subject and environment independent features extracted from the CSI data for training the classifiers. However, this approach requires a large amount of training data that must be collected from a lot of persons in different environments [15] to train a machine learning model. The other approach proposes the use of a semi-supervised learning technique, that requires users to manually label the activity fingerprints that may have been changed due to changes in the environment [39]. This solution requires user interaction that is not very practical for applications in elderly care.

A location and position invariant gesture recognition system “WiAG” was proposed in [35]. WiAG, first collects training samples from users using a single configuration. Next, it uses a translation function to virtually generate data samples for all possible configurations. Afterwards, it constructs classification models for each configuration using corresponding virtual samples. At runtime, WiAG first estimates the configuration of the user and then evaluates the runtime (or testing) data against the classification model corresponding to that configuration.

Wi-Motion [21] leverages amplitude and phase information extracted from the CSI data. At first, Wi-Motion separately constructs two support vector machine (SVM) classifiers using statistical and frequency domain features extracted from amplitude and phase information. To classify the activities, a posterior probability-based decision-level fusion strategy was employed to fuse the results of the classifiers. Reportedly, the Wi-Motion was able to achieve a recognition accuracy of 96.6% when evaluated using a data set consisting of five activities such as, bend, half squat, step, jumping, and stretching a leg.

In [25], the authors presented a spectrogram-based approach to compute environment independent fingerprints of different activities from the CSI data. At first, the impact of the environment and noise is removed from the CSI data. Afterwards, the spectrogram method was used to compute the mean Doppler shift (MDS) from the processed CSI data. The MDS represents the Doppler signature of the performed activity. Thereafter, different statistical and frequency domain features were extracted from the MDS.

They employed a SVM classifier to classify different activities based on the features extracted from the MDS. This approach achieved an overall recognition accuracy of 96.2% when evaluated using a data set consisting of four activities including walking, falling, sitting, and bending. As described above, the learning algorithms used in [21, 25] are able to classify human activities with overall recognition accuracies of 96.2% and 96.6%, respectively. However, these approaches [21, 25] require extensive user interaction and domain knowledge to extract and choose discriminative features from the processed CSI (or MDS) data, to effectively classify human activities.

1.2 Contributions and organization

This work presents Wi-Sense that uses a CNN to recognize human activities from environment-independent time-variant micro-Doppler signatures (or fingerprints) extracted from Wi-Fi CSI data. In contrast to previous works [7, 15, 40, 44], Wi-Sense uses both amplitude and phase information of the CSI data. First, Wi-Sense processes the CSI data to remove the impact of noise and fixed (or non-moving) objects present in the environment. The processed CSI data is then used to compute the spectrogram¹ of the CSI data corresponding to different human activities. These spectrograms represent the radio channel Doppler characteristics caused by fixed and moving objects present in the environment. As we know, the static objects do not cause any variation in the RF signals Doppler frequencies. This implies that different static objects’ positions in the environment will not influence the performance of our HAR system. The spectrograms are stored as portable network graphics (PNG) images and used to train a deep CNN. We evaluate this novel approach using a Wi-Fi CSI data set [25]. This data set were collected in an indoor environment from nine volunteers. Each volunteer carried out four different activities: walking, falling, sitting on a chair, and picking up an object from the floor.

In contrast to the approaches developed in [21, 25], the CNN used in Wi-Sense does not require extensive human interaction to manually extract features from the CSI data. The CNN used in the Wi-Sense is able to automatically extract discriminative features from the PNG-formatted spectrogram images and classify human activities. Moreover, best to our knowledge, this is the first work that provides an overview of HIS standards, their interoperability, and systematically describes how the proposed Wi-Fi-based HAR system can be integrated into the existing HIS standards.

¹A spectrogram provides a visual representation of the time-variant spectral distribution of a signal.

The rest of the paper is organized as follows. In Section 2, we provide a succinct overview of the proposed Wi-Sense HAR system. In Section 3, we present the details of the experimental setup and human activity data collection. Section 4 describes the steps involved in processing the CSI data and computing the spectrogram. In Section 5, we present the architecture of our CNN model, the classification process, and the obtained results. The details about the relevant HIS standards and the integration of the Wi-Sense HAR system into the HIS infrastructure are given in Section 6. Finally, Section 7 concludes this work and presents the future work.

2 Overview of the Wi-Sense system

The Wi-Sense HAR system comprises of three main modules, namely the RF sensing module, the data processing module, and the classification module. The RF sensing module of Wi-Sense consists of a Wi-Fi transmitter (T_x) and a Wi-Fi receiver (R_x). The T_x and R_x are Wi-Fi network interface cards (NICs) that operate in the 5 GHz band [14]. The T_x and R_x are used capture the CSI, while a user is performing different activities in an indoor environment. The T_x continuously emits RF signals that propagate in the indoor environment. While traveling from the T_x to the R_x , these RF signals reflect from the static and moving objects present in the environment, as shown in Fig. 1. The examples of static objects include walls, ceiling, and furniture, whereas the moving objects are the body segments of the moving person, such as feet, legs, hands, arms, head, and trunk.

The ambient RF signals experience frequency shift due to moving objects that are present in the environment. This phenomenon is known as the Doppler effect. The R_x receives these modified RF signals and reports the estimated CSI.

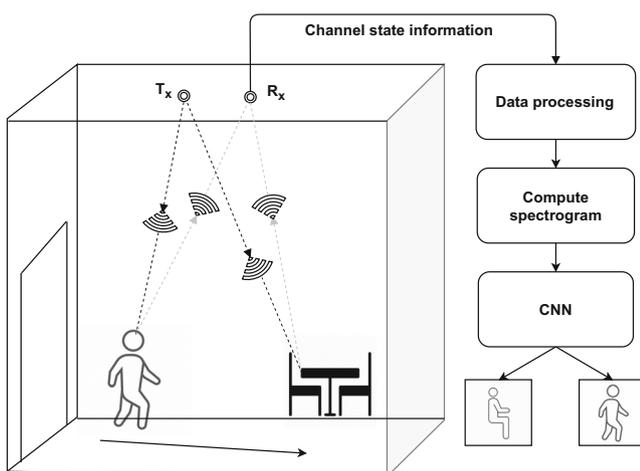


Fig. 1 The overview of the Wi-Sense HAR system

The Wi-Sense's data processing module first effectively reduces the noise from the CSI data. Thereafter, time-variant micro-Doppler signatures are extracted using the spectrogram method. The spectrogram images show the time-variant Doppler characteristics of the RF channel caused by the static and moving objects. As we know, the Doppler effect is only caused by the moving objects present in the environment, while scattered signal components received from static objects will not experience any Doppler shift. Therefore, we argue that variations in the micro-Doppler signatures are in fact due to the moving object, and thus, the performance of the Wi-Sense will not be affected by different static objects' placements. The spectrogram images are stored as PNG images and passed to the classification module. The Wi-Sense classification module is basically a CNN, which determines the types of activities performed by the user. In addition to that, the Wi-Sense HAR system can be integrated into the HIS infrastructure using a Tele-Care alarm service (see Section 6.2). Thus, upon detecting an accidental fall, it can send a fall alarm message to a health service provider.

3 Experimental setup and channel state information collection

In this paper, we used a human activity data set [25] that was collected in an indoor environment from nine volunteers in total. Each volunteer performed four different activities, namely walking, falling on a mattress, sitting on a chair, and picking up an object from the floor. During the data collection, only a single volunteer was moving inside the room. For each volunteer, we recorded several trials of each activity, and we asked the volunteer to stay inactive for one second before commencing and after finishing an activity trial.

Each volunteer repeated the walking activity 10 times by walking back and forth from points A to B and B to A, as shown in Fig. 2. The falling activity was carried out at point

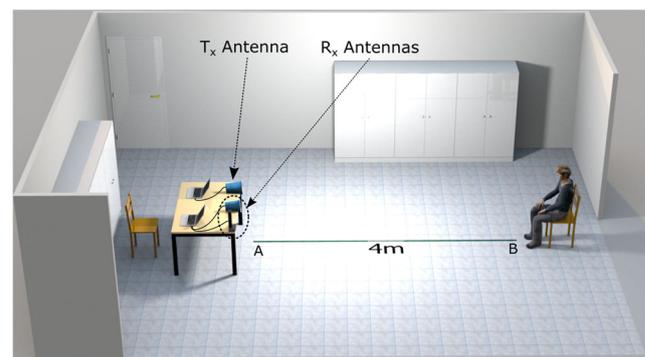


Fig. 2 The experimental setup for data collection [25]

B, where a thick mattress was placed on the floor and the participant fell on the mattress. Each volunteer repeated 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.

The sitting activity was also carried out at point B, where an arm-less chair was placed. Each volunteer repeated the sitting activity five times. For each sitting activity trial, the volunteer stood still next to the chair facing towards the antennas and then sat on the chair, as shown in Fig. 2. Finally, for the activity “Picking up an object from the floor,” a small object, e.g., a whiteboard marker, was placed on the floor at point B. We asked the volunteers to pick it up from a standing position. Each volunteer also repeated this activity five times.

To collect and parse the CSI data while the participants were performing the aforementioned activities, we used two laptops, each equipped with an Intel 5300 Wi-Fi NIC [14]. We installed the CSI Tool [10] on both laptops. One laptop was used as a transmitter (T_x), and the other laptop as a receiver (R_x). The internal antennas of the laptops normally have a limited coverage range; and therefore, we connected one external directional antenna to the T_x and two external antennas to the R_x , one of which was a directional and the other an omnidirectional antenna. The transmitting and receiving antennas were placed on a table as shown in Fig. 2 at a height of 0.8 m from the floor. The NICs of the T_x and R_x were operating at the 5.745 GHz central frequency with 20 MHz bandwidth in single-input multiple-output (SIMO) transmission mode. We used the “injector-monitor Wi-Fi mode,” where the T_x was set to transmit random data packets into the RF channel at a sampling frequency of 1 ms. The R_x was configured to receive the transmitted packets and report the estimated CSI in a matrix form for every received data packet. Generally, the dimension of the CSI data matrix is $N_{T_x} \times N_{R_x} \times K$, where N_{T_x} indicates the number of transmit antennas, N_{R_x} stands for the number of receive antennas, and K represents the number of OFDM subcarriers. Moreover, the CSI Tool reports estimated CSI data along 30 OFDM subcarriers for each transmission link. Therefore, in our case, the dimension of the CSI data matrix was $1 \times 2 \times 30$.

4 Processing of channel state information and spectrogram computation

The raw CSI data contains amplitude and phase information, which are corrupted by noise; and therefore, the raw CSI data streams cannot effectively be used to extract micro-Doppler signatures [24]. The CSI data amplitude is mainly corrupted by the ambient noise and adaptive changes of the transmission parameters [41]. In addition to that, the phase

of the CSI data suffers from errors introduced by the carrier frequency offset (CFO) and the sampling frequency offset (SFO) [37, 41]. The errors related to the CFO and SFO are due to the asynchronicity between the T_x and R_x clocks.

4.1 Phase correction

The first step towards extracting the micro-Doppler signatures from the CSI data requires CSI phase distortions elimination. In the literature, there exist three different methods that can be used to eliminate the phase distortions, namely the phase sanitization method [6], the phase calibration method with back-to-back channel configuration [18], and the CSI ratio method [42]. The phase sanitization method applies a linear transformation to the measured (distorted) phases to determine the true phase. However, it has been reported in [1] that the true phases obtained after applying the phase sanitization method do not unfold the Doppler features of the measured CSI data. The back-to-back channel configuration method [18] splits the transmitted signal into two similar signals using a two-way splitter, which is connected to the T_x . One of the two outputs of the splitter is connected to the T_x antenna, whereas the other output of the splitter is directly connected to one of the three RF antenna ports of the R_x using an RF cable, to set up a back-to-back channel. Moreover, the R_x antennas are connected to the remaining RF antenna ports of the R_x . In this way, the signal is first split and then it is simultaneously transmitted over the wireless and back-to-back channels. The R_x receives signals over both channels. As the signal received via the back-to-back channel has no phase distortions, it can be used to calibrate the phase of the signal received by the R_x antennas.

Inspired by the previous studies [24, 25, 42], we use the CSI ratio method [42] in this work. The CSI ratio method is more economical and simpler to set up, because it does not require additional hardware compared to the back-to-back channel configuration method [24]. Moreover, our experiments and previous works [24, 25] suggest that the CSI ratio method is more effective in reducing the phase distortions and noise from the amplitude information, compared to the phase sanitization and back-to-back channel configuration methods. In the CSI ratio method, at least one T_x and two R_x antennas are used. The R_x antennas are placed close to each other to simultaneously receive the CSI data. Thereafter, the CSI ratio $R(f'_k, t)$ is computed by dividing the CSI data of the first transmission link $H_{1,1}(f'_k, t)$ by the CSI data of the second transmission link $H_{1,2}(f'_k, t)$, i.e.,

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

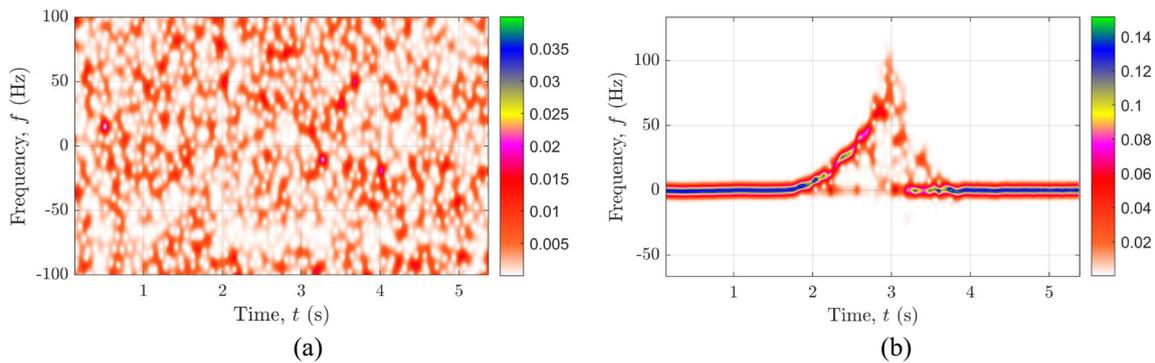


Fig. 3 The falling activity spectrograms obtained **a** without and **b** with applying the CSI ratio-based phase calibration technique

where $H_{i,j}(f'_k, t)$ is denotes the time-variant channel transfer function (CTF)² of a transmission link between i th transmit and j th receive antenna pair sampled at the k th subcarrier f'_k [24]. The used CSI Tool provides estimated CSI data for $K = 30$ OFDM subcarriers. Each subcarrier f'_k (where, $k = 1, 2, \dots, K$) can be expressed as

$$f'_k = f'_0 + k\Delta f' \quad (2)$$

where f'_0 represents the carrier frequency, k stands for the subcarrier index, and $\Delta f'$ indicates the subcarrier bandwidth. As the CSI Tool reports 30 CSI streams for each transmission link. Thus, we can obtain 30 CSI ratio streams after applying the CSI ratio method.

To demonstrate the effectiveness of the CSI ratio method in eliminating the CSI phase distortions, in Fig. 3, we present the comparison of the spectrograms of the CSI data that belong to a single trial of the falling activity before and after using the CSI ratio method. Note, the data processing steps explained in this subsection and in the following Sections 4.2 and 4.3 were kept the same for computing these spectrograms.

4.2 Dimensionality reduction

It has been reported in [37] that the variations of the RF signals due to human movements are correlated across different CSI data streams. With reference to Eq. 1, this means that CSI ratio streams $R(f'_k, t)$ corresponding to different subcarrier f'_k are correlated as well. Therefore, we apply the principal component analysis (PCA) [17] to remove correlated and redundant information from the CSI ratio streams. The PCA is a statistical method commonly applied to real value data sets to reduce the data dimensions and filter noise from the data. However, the CSI ratio streams are complex-valued. Therefore, we use the formulation of PCA applicable to the complex-domain, as described in [26].

²Within the scope of this work, the terms CSI and CTF are considered interchangeable.

The CSI ratio streams are indeed continuous time series. Therefore, before applying the PCA, the CSI ratio streams are arranged in the form of a matrix. This is done by considering the samples of $R(f'_k, t)$ at $t = t_n = nT$ for $n = 1, 2, \dots, N$, where T is the sampling interval and N denotes the number of samples in the time domain [24]. As a result, we obtain $R_{kn} = R(f'_k, t_n)$, which is used to express the CSI ratio streams in the form of a matrix. We call this matrix the CSI ratio matrix \mathbf{R} , and it is defined as follows

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1} & R_{K2} & \cdots & R_{KN} \end{pmatrix} \in \mathbb{C}^{K \times N}. \quad (3)$$

The CSI ratio matrix \mathbf{R} is a $K \times N$ complex matrix³. Each row of matrix \mathbf{R} simply represents the discrete form of the corresponding continuous CSI ratio stream. Next, the CSI ratio matrix \mathbf{R} is mean normalized by subtracting the mean value $\bar{m}_k = \sum_{n=1}^N R_{kn}/N$ from each row of the CSI ratio matrix \mathbf{R} , i.e.,

$$\mathbf{R}_m = \begin{pmatrix} R_{11} - \bar{m}_1 & R_{12} - \bar{m}_1 & \cdots & R_{1N} - \bar{m}_1 \\ R_{21} - \bar{m}_2 & R_{22} - \bar{m}_2 & \cdots & R_{2N} - \bar{m}_2 \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1} - \bar{m}_K & R_{K2} - \bar{m}_K & \cdots & R_{KN} - \bar{m}_K \end{pmatrix} \in \mathbb{C}^{K \times N}. \quad (4)$$

Thereafter, we compute the covariance matrix \mathbf{C} as follows

$$\mathbf{C} = \frac{1}{N-1} \mathbf{R}_m \mathbf{R}_m^H \quad (5)$$

where $(\cdot)^H$ denotes the conjugate transpose operator. The covariance matrix \mathbf{C} is a complex square (or Hermitian) matrix of dimension $K \times K$. The diagonal elements of \mathbf{C} describe the variances of the CSI ratio streams, and the off-diagonal elements describe their covariances [24]. Thereafter, we factorize the covariance matrix \mathbf{C} into eigenvectors and

³For the ease of notation, capital letters K and N are used to represent the dimensions of the CSI ratio matrix \mathbf{R} .

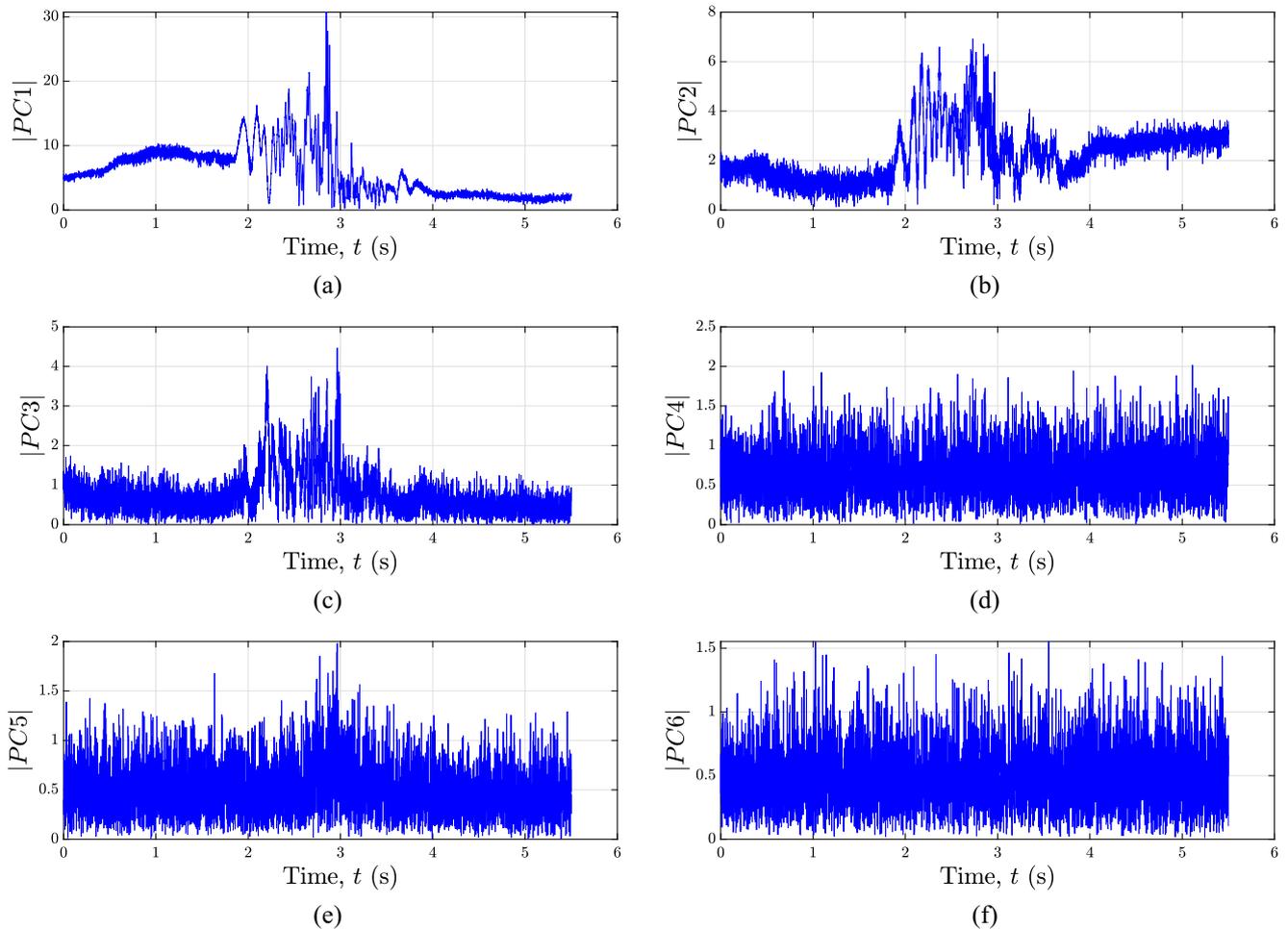


Fig. 4 The amplitude plots of the first six principal components of the falling activity

eigenvalues. Consequently, we obtain K complex-valued eigenvectors and K real-valued eigenvalues. The eigenvectors are sorted according to the decreasing eigenvalues. These eigenvectors are called the principal components or (principal axes). Let \mathbf{Z} be the matrix consisting of K eigenvectors sorted in order of decreasing eigenvalues. Thereafter, we project the mean normalized CSI ratio matrix \mathbf{R}_m onto these principal axes according to

$$\mathbf{Y} = \mathbf{R}_m^T \mathbf{Z} \tag{6}$$

where $(\cdot)^T$ is the transpose operator. The first principal component in \mathbf{Y} indicates the direction in which the data has the highest variance. On the contrary, the last principal component in \mathbf{Y} indicates the direction in which the data varies the least. So, the first principal component captures the maximum and the last principal component captures the minimum original data information. In Fig. 4, we visualize this effect by plotting the amplitude plots of the first six principal components of the falling activity.

In Fig. 4, we can observe that the first principal component (see Fig. 4a) captures the variations caused by the

falling activity, and it contains very less noise. Besides, we also observe that the level of noise increases in each subsequent principal component, as shown in Fig. 4. From the fourth onward till the last (Figs. 4c–f), each principal component is presents in the noise in the data. Therefore, we only used the first principal component for computing the spectrogram. Moreover, we apply a low-pass filter with a cut-off frequency of 150 Hz to the selected principal component to further minimize the effect of the high frequency components which are not caused by the human movement.

4.3 Computing the spectrogram

We use a two-step processor to compute the spectrogram of the filtered data $Y_1(t)$. In the first step, we compute the shorttime Fourier transform (STFT) of the filter data as follows

$$X_{Y_1}(f, t) = \int_{-\infty}^{\infty} Y_1(t')g(t' - t)e^{-j2\pi ft'} dt' \tag{7}$$

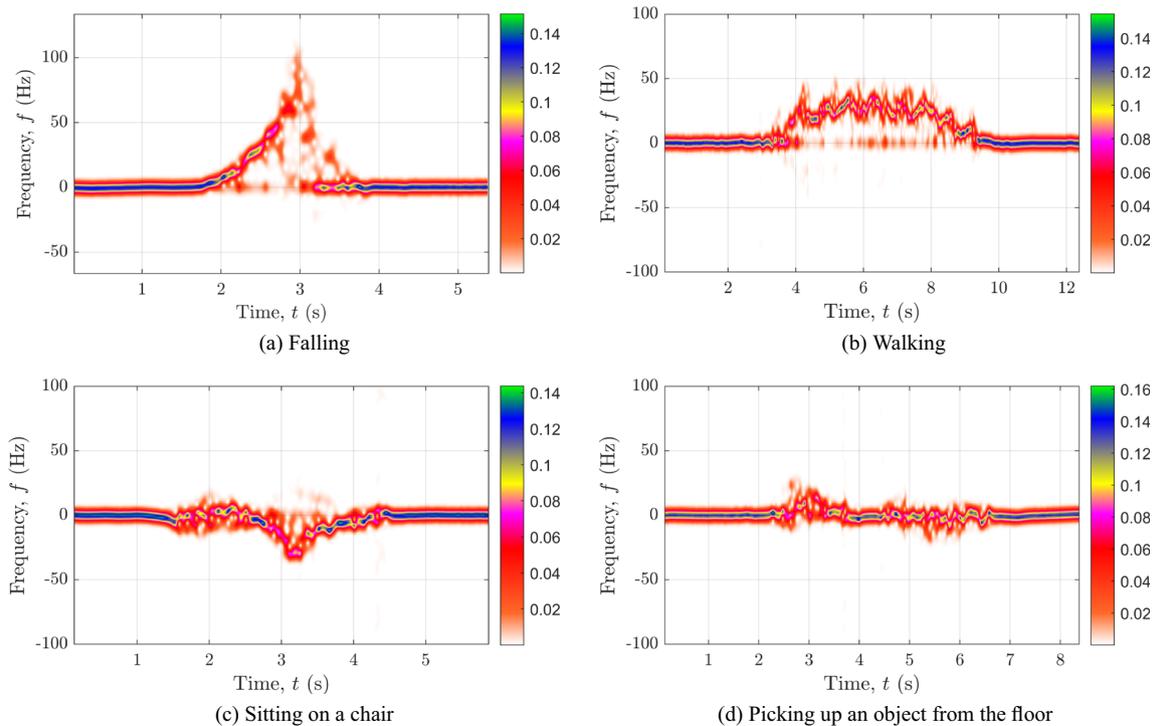


Fig. 5 Spectrograms of four activities explored in this work

where t' denotes the running time, t denotes the local time, and $g(t)$ is a Gaussian sliding window function, which is defined as

$$g(t) = \frac{1}{\sqrt{\sigma_w \sqrt{\pi}}} e^{-\frac{t^2}{2\sigma_w^2}}. \quad (8)$$

In Eq. 8, σ_w indicates the spread of the Gaussian window. The Gaussian window function $g(t)$ is real and has unit energy, i.e., $\int_{-\infty}^{\infty} g^2(t) dt = 1$. In the second step, the spectrogram $S_{Y_1}(f, t)$ is obtained as [4]

$$S_{Y_1}(f, t) = |X_{Y_1}(f, t)|^2. \quad (9)$$

The spectrograms of the activities that are explored in this work are shown in Fig. 5.

5 Classifying spectrogram images using convolutional neural network

For every activity trial in the collected data, we first computed the spectrogram and then saved it as a PNG image in a folder labelled with the activity. Thereafter, all spectrogram images were scaled to the same $224 \times 224 \times 3$ dimension by applying the bicubic interpolation technique [11]. We split the spectrogram image data set into the training, validation, and testing data sets. The training data set consist of 70% of the total data, whereas the remaining 30% data were divided equally into the validation and testing data sets. We

used the training data set to train the CNN with a batch size of 16. The architecture of our CNN is shown in Fig. 6. The CNN consists of 14 layers including input, flatten, and output layers. The dimensions (i.e., height and width) of the filters used in all convolutional layers are 5×5 and in all max-pooling layers 2×2 . The stride parameter, which defines the number of cell shifts over the given data matrix, was set to 1 for the convolutional layers and to 2 for the max-pooling layers. The number of filters used in the first, second, and third convolutional layer was 32, 48, and 64, respectively. All convolutional layers used the rectified linear unit (ReLU) activation function. After each max-pooling layer, a dropout layer (indicated by a green circle in Fig. 6) with a threshold 0.3 was used. The last two layers are fully connected with dimensions 256×1 and 84×1 , respectively. The dimension of the output layer is 4×1 and uses the Softmax activation function. The validation data were used to monitor the training progress of the CNN and to stop the training if the validation accuracy does not improve over 8 consecutive epochs. The CNN model accuracy and loss over the training and validation data are presented in Fig. 7a and b, respectively. Finally, the performance of the CNN model was evaluated based on the testing data set.

We use a confusion matrix to quantitatively visualize the results of our CNN model (see Fig. 8). In this confusion matrix, the counts of the correctly classified activity trials (i.e., true positive (TP) and true negative (TN)) are given in the green cells. The counts of the incorrectly classified

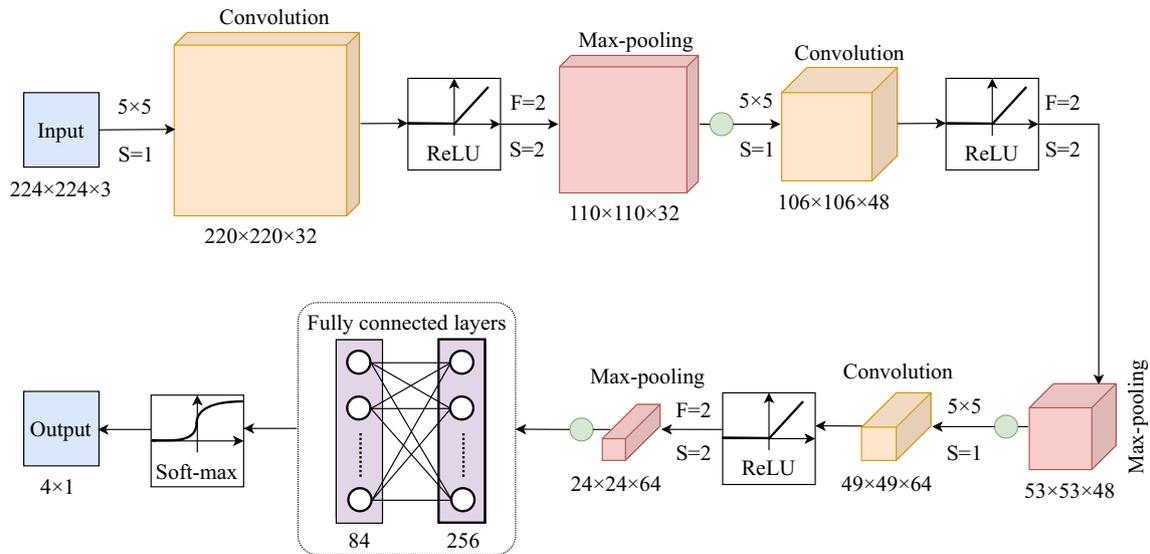


Fig. 6 The architecture of the CNN, where the symbol S indicates the stride parameter and the symbol F denotes the size of filter used in the max-pooling layer. The green circles represent the dropout layer

activity trials (i.e., false positive (*FP*) and false negative (*FN*)) are given in the red cells. Within the scope of this work, *TP* indicates the number of correctly classified activity trials of the positive class, and *TN* indicates the number of correctly classified activity trials of the negative class. On the other hand, *FP* represents the number of activity trials that actually belong to the negative class but incorrectly classified as the positive class. Similarly, *FN*

indicates the number of activity trials that actually belong to the positive class but incorrectly assigned to the negative class. The overall accuracy of the CNN model is given in the blue diagonal cell. The confusion matrix also presents the performance of the CNN model in terms of precision and recall for each class, which are presented in the rightmost column and the last row of the confusion matrix (see

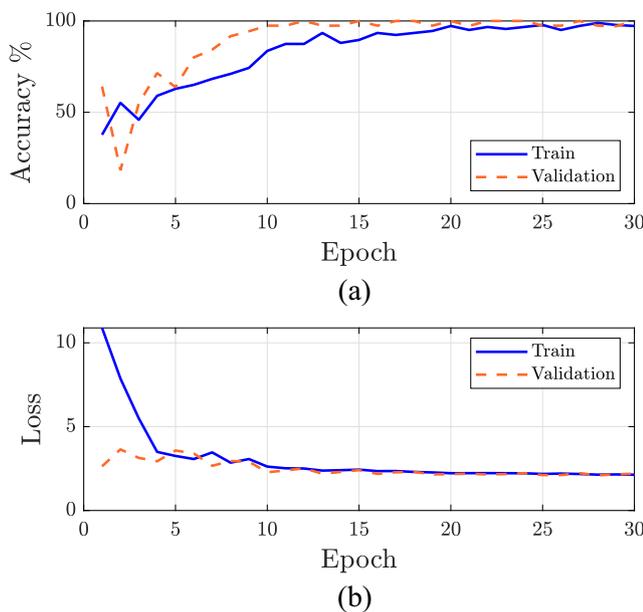


Fig. 7 **a** Accuracy and **b** loss during the training process based on the training and validation data sets

Predicted Class	Walk	15	0	0	0	100% 0.0%
	Fall	0	14	0	1	93.3% 6.7%
	Pick-object	0	0	8	0	100% 0.0%
	Sit	0	0	0	7	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	87.5% 12.5%	97.78% 2.22%
	Walk	Fall	Pick-object	Sit		
	Actual Class					

Fig. 8 Confusion matrix of the results obtained from the CNN model. The rightmost column and the last row indicate the precision and recall of the CNN model with respect to each class

Fig. 8), respectively. The expressions of precision, recall, and accuracy are given as

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (11)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \quad (12)$$

As shown in Fig. 8, the overall recognition accuracy of the CNN is 97.78%. The CNN makes a single classification error by wrongly predicting a sitting activity trial as the falling activity. The model precision for walking, falling, picking up an object, and sitting activity is 100%, 93.3%, 100%, and 100%, respectively. The recall of the sitting activity is 87.5%, whereas the other three activities have a recall of 100%.

Moreover, Wi-Sense recognizes the activities performed at greater distances. For instance, three out of the four activities are performed at a distance of 4 m from the transmitting and receiving antennas.

It is not possible to directly compare the results of Wi-Sense with existing works presented in Section 1.1, because these approaches [7, 21, 44] use different data sets that contain different activities. Besides, the data sets used in these approaches are also not publicly available. However, we can compare the results of Wi-Sense with [25], because the same data set were used to evaluate the performance of Wi-Sense and the approach presented in [25]. By comparing the results of Wi-Sense with the results of [25] as described in Section 1.1, we can conclude that Wi-Sense performs better than the approach developed in [25].

6 HAR in the context of health information systems

In this section, we briefly describe how HAR systems and eHealth technologies contribute in developing adaptable and safe environments for elderly and patients to support them to enjoy independent living in their homes. In Section 6.1, we first provide a succinct overview of existing HIS standards that are relevant to HAR and then in the following Section 6.2, we describe in detail how the proposed Wi-Sense HAR system can be integrated into the HIS infrastructure.

The development of telehealth and telecare technologies as well as solutions for the monitoring of ageing citizens and patients at home is one of the most dynamic areas within the eHealth domain. Telehealth solutions mostly allow automatic and seamless monitoring of various factors of the patients' health condition using wearable sensor devices attached to the patients' bodies. In addition, digital

assessment tools and questionnaires are used for interactive assessment of more subjective healthcare and wellness symptoms. Community telecare solutions cover social care support (including emergency alarms), dementia care, and assisted living by utilizing HAR with different types of in-house surveillance technologies. One of the main objectives of in-home monitoring systems is the detection of critical health conditions. Typical cases are a temporary unconsciousness due to a general collapse, a stroke, a heart attack, or a fall (often in conjunction with a shock or serious injuries). In such situations, it is desirable that the functions of the HAR module include the following: identification of abnormalities from normal activities and behaviour, such as fall events; assessment and analysis of the risk condition, such as the patient is inactive for a certain amount of time after the fall event; and triggering a corresponding alarm.

The in-home monitoring system has to initiate a corresponding support request. This can be the automatic initiation of an emergency phone call, the creation of a short message service, or an instant message to a formal or informal caretaker. As part of a telecare system, the in-home monitoring system can forward specific electronic messages to a dedicated social alarm/emergency service within the HIS infrastructure, including additional information about the type of alarm and the person's condition.

6.1 Interoperability

In-home activity monitors, fall detection, and other personal eHealth solutions, such as vital-sign monitoring mostly involve multi-vendor devices. Therefore, such eHealth solutions mainly face interoperability challenges [28]. According to the Institute of Electrical and Electronics Engineers (IEEE), interoperability in eHealth solutions is defined as “the ability of two or more systems or components to exchange information and use the information that has been exchanged” [9, 28]. This means the devices and standards followed to develop eHealth solutions must be able to interact with each other by using a common language consisting of the common naming conventions, data types, data formats, message syntax, data encoding, and data decoding rules [28]. In addition to interoperability, adhering to the common language is essential for developing scalable eHealth solutions [5].

A number of non-profit organizations, such as Integrating the Healthcare Enterprise (IHE)⁴ and the Personal Connected Health Alliance (PCHA)⁵ work towards the development of widely accepted standards for eHealth interoperability. IHE addresses the interoperability of eHealth systems. They promote the adoption of the Continua Health Alliance

⁴<https://www.ihe.net/>

⁵<https://www.pchalliance.org/about-continua>

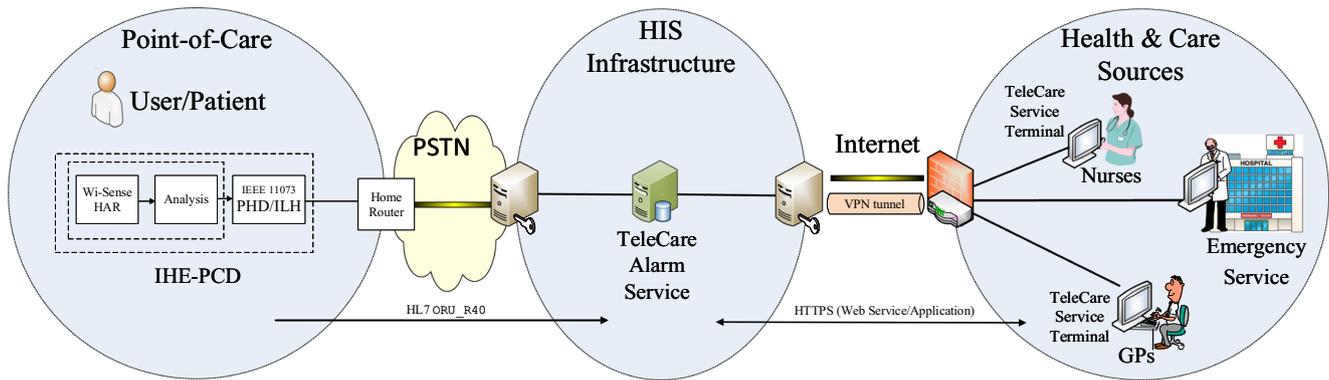


Fig. 9 Integration of the Wi-Sense system into the HIS infrastructure

(ISO/IEEE 11073) specifications in the patient care device (PCD) domain and a Health Level Seven International (HL7)⁶ standard for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice, covering in particular clinical systems within the HIS infrastructure [30].

PCHA regularly publishes the Continua design guidelines that define the framework of underlying standards and criteria to meet interoperability of components for applications used for monitoring personal health and wellness [27]. ISO/IEEE 11073:10471 is one of the ISO/IEEE 11073 family of standards that define the common core of communication functionality for independent living hubs (ILHs). Within the scope of ISO/IEEE 11073:10471 standard, the ILH is a device that is responsible for communicating with binary sensors, normalizing the information obtained from these sensors, and forwarding this information to one or multiple managers. In this context, the binary sensors are also known as situation monitors, such as smoke sensors, fall sensors, motion sensor, door sensors, enuresis sensors, and chair or bed occupancy sensors. The information obtained from the situation monitors can be examined when a person's activities have deviated significantly from the normal behavior.

6.2 Integration of Wi-Sense in the health information system infrastructure

Utilizing the standards noted in the previous subsection, and considering the telehealth and telecare platform for remote monitoring as proposed in [20], we propose to integrate the Wi-Sense HAR system using the TeleCare alarm service within the HIS infrastructure, as illustrated in Fig. 9. The key elements in Fig. 9 are point-of-care, HIS infrastructure, and health and care sources. As shown in Fig. 9, the ILH and situation monitoring sensors are deployed at

the point-of-care (or user side). The ILH processes and normalizes the information obtained from sensors and generate messages/alarm following an alarm communication management profile [13]. The ILH is connected to the TeleCare alarm service in HIS infrastructure using the public switched telephone network (PSTN). The TeleCare alarm service forwards the messages to the health and care sources using a virtual private network.

As described in Section 2, the Wi-Sense HAR system consists of RF sensing, data processing, and classification modules. Within the context of this work, the Wi-Sense's RF sensing module acts as a motion and fall sensor, following the Continua/IEEE 11073:10471 specifications. The data processing and classification modules of the Wi-Sense process and classify the information obtained from the RF sensing module to detect accidental falls and other types of activities. Therefore, the Wi-Sense HAR system is realized as an ILH, which can be deployed at the point-of-care. This ILH follows the IHE-PCD-04 Alarm Communication Management profile. The IHE-PCD-04 profile defines an HL7 ORU_R40 message for communicating alerts that requires a timely response from the health/emergency service provider [13]. Therefore, the Wi-Sense HAR system uses an HL7 ORU_R40 message to communicate fall alarm to the TeleCare alarm service in the HIS infrastructure, as shown in Fig. 9. The underlying format of the HL7 ORU_R40 message can be found in [13]. The TeleCare alarm service can be realized as a web application server that forwards the fall alarm using the HTTP/HTTPS messages to the dedicated emergency service providers as well as to the TeleCare support service provider.

7 Conclusion and future work

In this work, we presented the Wi-Sense HAR system, which combines RF sensing and deep learning techniques to recognize human activities including falls. In addition,

⁶<http://www.hl7.org/>

we also provided an overview of the existing relevant HIS standards and explained how the proposed Wi-Sense HAR system can be realized according to these standards.

The sensing module of Wi-Sense uses two laptops, where one laptop acts as a transmitter and the other as a receiver to collect the CSI data. We collected CSI data while nine participants performed four activities, namely walking, falling on the mattress, sitting on a chair, and picking up an object from the floor. A three-step process was used to filter the collected CSI data. At first, we applied the CSI ratio method to the collected CSI data to reduce the impact of the phase offset. In the subsequent step, the PCA was applied to remove redundant information from the data. In the last step, a low pass filter was used to reduce the impact of high-frequency components that were not caused by human movements. Thereafter, we computed a spectrogram for each activity trial in the collected data. These spectrogram images data set was divided into training, validation, and testing data sets. We used the training and validation data sets to train a 14-layer CNN and monitor the training process, respectively. The testing data set was used to evaluate the performance of the CNN. The results have shown that our CNN model achieved an overall accuracy of 97.78%. The Wi-Sense HAR system acts as an ILH following the Continua/IEEE 11073:10471 specifications. Upon detecting a fall, ILH generates an HL7 ORU_R40 message following the IHE-PCD-04 Alarm Communication Management profile and sends this message to the TeleCare alarm service, which forwards this message to the health service providers.

In the future, we will conduct more experiments to quantitatively evaluate the performance of our approach in different environments and develop a prototype of Wi-Sense system for real-time HAR and fall detection. Besides, we will practically demonstrate the integration of Wi-Sense in the HIS infrastructure. The Wi-Sense HAR system is limited to recognize activities of a single user. In the future, we extend Wi-Sense to recognize multi-person activities by employing multiple input multiple output (MIMO) antenna configuration.

Funding Open access funding provided by University of Agder. This work has been carried out within the scope of the WiCare project funded by the Research Council of Norway under the grant number 261895/F20.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abdelgawwad A, Borhani A, Pätzold M (2020) Modelling, analysis, and simulation of the micro-Doppler effect in wideband indoor channels with confirmation through pendulum experiments. *Sensors* 20(4):1049
2. Atif M, Muralidharan S, Ko H, Yoo B (2020) Wi-ESP—a tool for CSI-based device-free Wi-Fi sensing (DFWS). *J Computat Design Eng* 7(5):644–656
3. Beddiar DR, Nini B, Sabokrou M, Hadid A (2020) Vision-based human activity recognition: a survey. *Multimed Tools Appl* 79(41):30509–30555
4. Boashash B (2015) Time-frequency signal analysis and processing – a comprehensive reference, 2nd edn. Elsevier Academic Press, Cambridge
5. Bouamrane MM, Tao C, Sarkar I (2015) Managing interoperability and complexity in health systems. *Methods Inform Med* 54(1):1–4
6. Chen J, Li F, Chen H, Yang S, Wang Y (2019) Dynamic gesture recognition using wireless signals with less disturbance. *Pers Ubiquit Comput* 23(1):17–27
7. Chen Z, Zhang L, Jiang C, Cao Z, Cui W (2018) Wifi CSI based passive human activity recognition using attention based BLSTM. *IEEE Trans Mob Comput* 18(11):2714–2724
8. Ding J, Wang Y (2019) Wifi CSI-based human activity recognition using deep recurrent neural network. *IEEE Access* 7:174257–174269
9. Dixon BE, Rahurkar S, Apathy NC (2020) Interoperability and health information exchange for public health. Springer International Publishing, Cham, pp 307–324
10. Halperin D, Hu W, Sheth A, Wetherall D (2011) Tool release: gathering 802.11N traces with channel state information. *ACM SIGCOMM Comput Commun Rev* 41(1):53–53
11. Han D (2013) Comparison of commonly used image interpolation methods. In: 2nd international conference on computer science and electronics engineering. Atlantis Press, Atlantis, pp 1556–1559
12. Hsieh CF, Chen YC, Hsieh CY, Ku ML (2020) Device-free indoor human activity recognition using Wi-Fi RSSI: machine learning approaches. In: 2020 IEEE International conference on consumer electronics-Taiwan (ICCE-Taiwan). IEEE, pp 1–2
13. Integrating the Healthcare Enterprise (2019) IHE Patient care device technical framework, (IHE PCD F-2): Transactions. https://www.ihe.net/uploadedfiles/documents/PCD/IHE_PCD_TF_vol2.pdf Accessed 13.01.2021
14. Intel® (2020) Product Brief Intel Ultimate N WiFi Link 5300. Tech. rep., Intel Corporation, USA. <https://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/ultimate-n-wifi-link-5300-brief.pdf>. Accessed. 13.01.2020
15. Jiang W et al (2018) Towards environment independent device free human activity recognition. In: 24Th annual international conference on mobile computing and networking. ACM, pp 289–304
16. Jokanović B, Amin M (2018) Fall detection using deep learning in range-doppler radars. *IEEE Trans Aerosp Electron Syst* 54(1):180–189

17. Jolliffe I (2002) Principal component analysis. Springer, New York
18. Keerativoranan N, Haniz A, Saito K, Takada JI (2018) Mitigation of CSI temporal phase rotation with B2B calibration method for fine-grained motion detection analysis on commodity Wi-Fi devices. *Sensors* 18(11):1–18
19. Kim K, Jalal A, Mahmood M (2019) Vision-based human activity recognition system using depth silhouettes: a smart home system for monitoring the residents. *J Electric Eng Technol* 14(6):2567–2573
20. Lamprinakos G et al (2015) An integrated remote monitoring platform towards telehealth and telecare services interoperability. *Inf Sci* 308:23–37
21. Li H, He X, Chen X, Fang Y, Fang Q (2019) Wi-motion: a robust human activity recognition using WiFi signals. *IEEE Access* 7:153287–153299
22. Li X, He Y, Jing X (2019) A survey of deep learning-based human activity recognition in radar. *Remote Sens* 11(9):1–22
23. Loncar-Turukalo T, Zdravovski E, Machado da Silva J, Chouvarda I, Trajkovic V (2019) Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. *J Med Internet Res* 21(9):e14017
24. Maaaz M, Chelli A, Abdelgawwad AA, Mallofré AC, Pätzold M (2020) WiWeHAR: Multimodal human activity recognition using Wi-Fi and wearable sensing modalities. *IEEE Access* 8:164453–164470
25. Maaaz M, Chelli A, Pätzold M (2020) WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition. In: 2020 IEEE 91st vehicular technology conference (VTC2020-spring), pp 1–7
26. Papaioannou AD (2017) Component analysis of complex-valued data for machine learning and computer vision tasks. Ph.D. thesis Imperial College London
27. Personal Connected Health Alliance (2019) Continua design guidelines: h.810 introduction – interoperability design guidelines for personal connected health systems. <https://members.pchalliance.org/document/dl/2148>. Accessed 13.01.2021
28. Schmitt L, Falck T, Wartena F, Simons D (2007) Novel ISO/IEEE 11073 standards for personal telehealth systems interoperability. In: 2007 Joint workshop on high confidence medical devices, software, and systems and medical device plug-and-play interoperability (HCMDSS-MDPnp 2007). IEEE, pp 146–148
29. Sigg S, Scholz M, Shi S, Ji Y, Beigl M (2014) RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals. *IEEE Trans Mob Comput* 13(4):907–920
30. Sloane EB, Thalassinidis A, Silva R (2018) ISO/IEEE 11073, IHE, and HL7: fostering standards-based safe, reliable, secure and interoperable biomedical technologies. In: Southeastcon 2018. IEEE, pp 1–3
31. Smith KA, Csech C, Murdoch D, Shaker G (2018) Gesture recognition using mm-Wave sensor for human-car interface. *IEEE Sensors Lett* 2(2):1–4
32. Uddin MZ, Hassan MM, Alsanad A, Savaglio C (2020) A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. *Inform Fusion* 55:105–115
33. United Nations (2002) Madrid international plan of action on ageing. Tech. rep., Second World Assembly on Ageing, New York. https://www.un.org/en/events/pastevents/pdfs/Madrid_plan.pdf. Accessed 06.07.2020
34. United Nations (2019) World population ageing 2019: highlights. Tech. rep., Department of Economic and Social Affairs Population Division, New York. <https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf>. Accessed 06.07.2020
35. Virmani A, Shahzad M (2017) Position and orientation agnostic gesture recognition using WiFi. In: Proceedings of the 15th Annual international conference on mobile systems, applications, and services, pp 252–264
36. Wang M, Zhang YD, Cui G (2019) Human motion recognition exploiting radar with stacked recurrent neural network. *Digit Signal Process* 87:125–131
37. Wang W et al (2017) Device-free human activity recognition using commercial WiFi devices. *IEEE J Select Areas Commun* 35(5):1118–1131
38. Wang Y, Cang S, Yu H (2019) A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst Appl* 137:167–190
39. Wang Y et al (2014) E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures. In: 20th annual international conference on mobile computing and networking, pp 617–628
40. Yan H, Zhang Y, Wang Y, Xu K (2020) Wiact: a passive WiFi-based human activity recognition system. *IEEE Sensors J* 20(1):296–305
41. Yousefi S, Narui H, Dayal S, Ermon S, Valaee S (2017) A survey on behavior recognition using WiFi channel state information. *IEEE Commun Mag* 55(10):98–104
42. Zeng Y et al (2019) Farsense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 3(3):1–26
43. Zhang S et al (2017) A review on human activity recognition using vision-based method. *J Healthcare Eng* 2017:1–31
44. Zou H et al (2018) Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In: 2018 IEEE International conference on communications (ICC). IEEE, pp 1–6

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