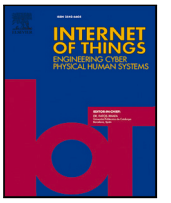
Internet of Things 24 (2023) 100960

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.elsevier.com/locate/iot



Review article

Channel phase processing in wireless networks for human activity recognition

Guillermo Diaz [a,*], Iker Sobron [b], Iñaki Eizmendi [a], Iratxe Landa [a], Johana Coyote [c], Manuel Velez [a]

[a] Department Communications Engineering, University of the Basque Country (UPV/EHU), Bilbao, 48013, Basque Country, Spain
[b] Department of Computer Languages and Systems, University of the Basque Country (UPV/EHU), Bilbao, 48013, Basque Country, Spain
[c] Department of Telecommunications, National Autonomous University of Mexico (UNAM), Mexico City, 04510, Mexico

ARTICLE INFO

Keywords:
CSI
Channel state information
Channel phase
Wireless sensing
Human activity recognition
HAR
Savitzky–Golay
Phase sanitization

ABSTRACT

The phase of the channel state information (CSI) is underutilized as a source of information in wireless sensing due to its sensitivity to synchronization errors of the signal reception. A linear transformation of the phase is commonly applied to correct linear offsets and, in a few cases, some filtering in time or frequency is carried out to smooth the data. This paper presents a novel processing method of the CSI phase to improve the accuracy of human activity recognition (HAR) in indoor environments. This new method, coined Time Smoothing and Frequency Rebuild (TSFR), consists of performing a CSI phase sanitization method to remove phase impairments based on a linear regression transformation method, then a time domain filtering stage with a Savitzky–Golay (SG) filter for denoising purposes and, finally, the phase is rebuilt, eliminating distortions in frequency caused by SG filtering. The TSFR method has been tested on five datasets obtained from experimental measurements, using three different deep learning algorithms, and compared against five other types of CSI phase processing. The results show an accuracy improvement using TSFR in all the cases. Concretely, accuracy performance higher than 90% in most of the studied scenarios has been achieved with the proposed solution. In few-shot learning strategies, TSFR outperforms the state-of-the-art performance from 35% to 85%.

1. Introduction

Wireless Sensing has been a rapidly growing field of study within the Internet of Things in recent years. It involves measuring wireless channel characteristics using existing wireless networks, such as WiFi networks, to sense environmental changes in the surrounding area of the network. Human activity recognition (HAR) in indoor environments is one of the main fields of application of wireless sensing.

The pervasive deployment of wireless networks worldwide and the fact that wireless sensing can be considered a privacy-preserving solution make this technology a promising alternative to other sensing solutions such as video surveillance with depth cameras or wearables [1–8]. Those other sensing methods present some drawbacks; for instance, cameras can compromise user privacy and, in the case of wearables, users should carry the devices on their person to be monitored. In wireless systems based on orthogonal frequency-division multiplexing (OFDM), such as WiFi, the Received Signal Strength Indicator (RSSI) and the Channel State Information (CSI) are used for wireless sensing. RSSI suffers from significant uncertainties due to the signal fluctuations under

* Corresponding author.
E-mail address: guillermo.diaz@ehu.eus (G. Diaz).

https://doi.org/10.1016/j.iot.2023.100960
Received 19 July 2023; Received in revised form 18 September 2023; Accepted 3 October 2023
Available online 12 October 2023

actual conditions, such as scattering, degradation, and sensitivity to noise [9]. Therefore, in recent years, CSI data have been widely used due to its major robustness against noise and other impairments of the signal reception. A time series of CSI measurements show how wireless signals propagate through objects and humans in the time, frequency, and spatial domains and can be used for different monitoring applications. Due to this, human activity recognition is an important field of wireless sensing, ranging in several areas such as crowd sensing [10–12], people localization [13–16], vital sign detection [17–19], and gesture recognition [20–28]. Likewise, CSI-based sensing can also be employed in other applications, such as electrical device classification based on the effect of the impulsive noise in the received signals [29]. In addition, it is worth noticing that IEEE 802.11 has recently approved a new task group named IEEE 802.11bf to accommodate sensing operations [30] into the WiFi standards.

Most of these papers use deep learning (DL) algorithms to make their predictions. In particular, fully-connected neural networks (FNN) and convolutional neural networks (CNN) are two of the most commonly deep neural network (DNN) models, which are also used in this work. As examples, FNNs have been used in works about health [31], credit risk [32], manufacturing [33] or recommendation systems [34], among others. For its part, CNNs have given very positive results in image analysis in different fields, such as health [35,36], aerospace industry [37], autonomous driving [38], agriculture and crops [39,40], robotics [41,42] or human resource recruiment [43], among others.

CSI measures the channel frequency response (CFR) of a wireless communication link based on OFDM. Several studies have explored channel modeling and characterization to optimize power allocation while minimizing interference among subcarriers [44–46]. In the context of our research, the power allocation methods inherent to the 802.11n standard suffice to operate under the assumption of a stable and efficient network. Utilizing CSI data, we obtain information in terms of both the amplitude and phase of the propagation channel for each subcarrier within an OFDM symbol. While the CSI amplitude provides a reasonably accurate estimation of the CFR amplitude, the phase contains uncertainties that make it challenging to use in many applications and theoretical developments of HAR. Correcting these phase uncertainties in frequency and time domains is a complex task, so many proposals in wireless sensing choose to work exclusively with amplitude [10,11,13–29].

However, some authors have developed methods to perform channel phase estimation of WiFi CSI for indoor monitoring. A wide variety of these techniques perform a linear transformation (LT) of the phase to correct the linear impairments caused by synchronization issues. In this regard, LT is first used by Souvik Sen et al. [47], where CSI is used to detect the position of people within different rooms, obtaining reasonable results. Qian, K. et al. [48] derive meaningful phase information by employing LT on the raw CSI to eliminate the significant random noise in the frequency domain. Outlier filtering is applied to shift out biased observations. Extracting various statistical features, such as variance, mean, and distribution distance, they obtain an accuracy of 90% for human motion when using three antennas. Wang, X. et al. [49] use LT method for correcting the phase and then employ a DNN with three hidden layers to train the calibrated phases. Their results for detecting human positions in two rooms have about a 20% of error in distance. Also, Fang, S. et al. [50] use LT to calibrate CSI phases, then an algorithm to extract different features is used, and last, a DNN classifies among three different human activities inside a car. In their work, Dang, X. et al. [51] perform phase LT before using the difference between adjacent subcarriers to train a backpropagation neural network with fingerprint data. Recently, Cheng, X. et al. [52] constructed the phase difference matrix expanded by the mean and standard deviation of the phase difference as a feature matrix after the LT method. Then Savitzky–Golay filter is performed on the raw CSI phase information. More recently, Bu, Q. et al. [53] introduce TransferSense, a one-time, environment-independent WiFi sensing method based on DL that converts RF sensing tasks into image classification and uses amplitude and phase data corrected with the LT method.

In addition to the LT method, other variants try to correct the errors of the estimated CSI phase, including non-linear errors. In the work of Kotaru, M. et al. [54], a similar linear phase calibration method is developed as an extension of the LT processing to multiple antennas, using the frequency difference between subcarriers to estimate the phase. Zhu, H. et al. [55] correct linear errors in phase, assuming that the CFR for one specific frequency should be the same even when measured in different bands. Then, they determine the time and frequency offsets in each band by matching the terms which define the CFR in each band. In Tadayon, N. et al. paper [56], the authors estimate time and frequency phase offsets separately. For time offsets, they assume that the channel impulse response is a linear combination of periodic functions whose period varies smoothly from sample to sample and try to correct the jumps observed at the power delay profile. For frequency offsets, the authors prove that the phase of the signal at the receiver follows a normal distribution to obtain an average value of the phase for all subcarriers in each packet or symbol. In a novel work, Meneghello, F. et al. [57], the authors consider a multipath propagation model for each CSI sample, using the most robust path as a reference to correct for time offsets in the remaining paths. They define each CSI sample as a product between the contributions that depend on the subcarriers index and a vector representing the independent terms from multipath. Then, for each subcarrier, they calculate the terms of that product and multiply it by the conjugate of the one with the strongest path to eliminate constant offsets in frequency.

After correcting the phase of the CSI, some works employ filtering to remove noise from the signal. One of the most recently used is the Savitzky–Golay (SG) filter, since it allows data smoothing with a reduced distortion of the signal tendency. This filter has been applied for wireless sensing in the frequency domain [52] and in the time domain [58,59].

In our previous work [60], a channel phase calibration method was presented based on a linear regression of the CSI phase. In addition, time smoothing of the phase was carried out through a SG filter, and finally, an algorithm was proposed to correct phase gaps in frequency. The calibration method was tested using the power profile of simulated wireless channels. Based on the previous work, this manuscript presents an improved and extended method of phase processing for HAR in wireless sensing. Additionally, to validate the proposed method, a comprehensive analysis of the proposal is performed over five different datasets of experimental measurements of HAR, using three different neural networks and comparing it with five other types of phase processing.

Considering the above, this paper focuses on channel phase processing to improve the accuracy of HAR classification in indoor environments with OFDM-based wireless signals. The contributions of this manuscript to the current state of the art are the following:

- We propose a novel phase processing method of CSI, coined Time Smoothing and Frequency Rebuild (TSFR), to be used for HAR. It consists, first, of an improved model for channel phase sanitization [60], adjusting and removing some parameters from the previous work. In addition, a new algorithm has been developed to smooth the phase in the time domain and correct discontinuities generated in frequency after filtering.
- Two new CSI-based datasets with real measurements have been generated for counting people and position localization in indoor environments.
- Two regular DNN have been designed for CSI-based HAR: a fully connected network with four hidden layers and one dropout layer, and a convolutional network with three convolutional layers, three max pool layers and two flatten layers. In addition, the few-shot meta-learning technique named ProtoNet [61] is also implemented to check the transferability of the results.
- It presents a comprehensive performance analysis of the TSFR proposal for HAR purposes over five datasets (two new, three from the bibliography) and three DL models (two new, one from the bibliography). In this analysis, the use of the SG filter in the time domain, frequency domain, and in both domains simultaneously has been assessed. In addition, performance comparisons have been carried out with the other methods from the state of the art. Furthermore, performance results in terms of accuracy and confusion matrices have been obtained when working with the processed CSI phase, the CSI amplitude, and both variables combined.

The rest of the paper is organized as follows: Section 2 summarizes the main concepts on which our proposal is based: CSI and the Savitzky–Golay filter. Section 3 presents the proposed method. The datasets and the DL algorithms which are utilized in this work are described in Section 4. The results and discussion are presented in Section 5. Finally, the paper will be concluded with some ideas and future directions in Section 6.

Notation: Matrices are represented in capital letters and boldface. The matrix $\boldsymbol{O}_{SxK}$ represents a zero matrix with S rows and K columns. $\boldsymbol{O}_{*,k}$ represents the column vector k, and $\boldsymbol{O}_{s,*}$ the row column s. The application of the Savitzky–Golay filter is represented as $SG\{\cdot, n, 2l+1\}$ being n the order of the polynomial used to fit the samples and l the length of the filter window.

2. Preliminary concepts

2.1. Channel state information

CSI describes the properties of the channel through which the signal propagates, in this case, OFDM wireless signals. These channel properties depend on the environment and the propagation medium and can therefore be used to extract characteristics of the environment. In the field of HAR, CSI is widely used because the channel properties are affected by environmental changes. So these variations are associated with the different activities to be classified.

For an OFDM system, the received signal in the frequency domain can be modeled as

$$\boldsymbol{y} = \boldsymbol{H} \cdot \boldsymbol{x} + \boldsymbol{z} \tag{1}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ denote the received and transmitted signal vectors, respectively, $\boldsymbol{z}$ is the additive complex white Gaussian noise, and $\boldsymbol{H}$ represents a diagonal matrix of the CFR, also referred as CSI. The CSI of the kth subcarrier during the sth symbol, $h_{s,k}$, is a complex value as follows:

$$h_{s,k} = |h_{s,k}|\, e^{j\theta_{s,k}} \tag{2}$$

where $|h_{s,k}|$ and $\theta_{s,k}$ are the amplitude and the phase, respectively. The CSI is therefore composed of two independent sources of information, amplitude on the one hand and phase on the other.

At the receiver side, CSI is usually estimated to decode the received signal. In this process, synchronization issues can lead to several errors in the estimated CSI, making the treatment of the phase complex for HAR purposes due to its uncertainties and offsets. In particular, there are three main types of errors [62] affecting the phase that do not reduce communication quality but are of great importance when working with CSI for HAR classification in closed environments.

- Sample Frequency Offset (SFO) is due to a mismatch of the oscillators between the transmitter (TX) and the receiver (RX). This lack of synchronization generates a time shift of the received signal concerning the transmitted signal. As the local oscillator remains stable over a short time, the SFO is usually treated as a constant.
- Sample Time Offset (STO) occurs because the receiver detects the packet by correlation operation and signal power calculation. Due to hardware imperfection, this process introduces a random time shift.
- Carrier Frequency Offset (CFO) occurs because the receiver center frequency is not synchronized. The system completes the estimation and compensation at the receiver by analyzing the cyclic prefix and pilot signals. However, due to hardware instability, the frequency offset cannot be entirely determined, and this residual offset causes a non-negligible error in the phase.

Therefore, let $\widehat{\boldsymbol{H}}_{SxK}$ be the estimated CSI matrix of S symbols with K subcarriers:

$$\widehat{\boldsymbol{H}}_{SxK} = \begin{pmatrix} \hat{h}_{1,1} & \cdots & \hat{h}_{1,K} \\ \vdots & \ddots & \vdots \\ \hat{h}_{S,1} & \cdots & \hat{h}_{S,K} \end{pmatrix} \tag{3}$$

where the (s,k)th element of $\widehat{\boldsymbol{H}}$ can be given by $\hat{h}_{s,k} = \left|\hat{h}_{s,k}\right| e^{j\hat{\theta}_{s,k}}$. Note that each row in Eq. (3) corresponds to one diagonal in H from Eq. (1). Likewise, a matrix of the measured CSI phases can be defined as $\hat{\boldsymbol{\Theta}}_{S\times K}$ where the measured phase at the kth subcarrier of the sth CSI frame can be expressed as:

$$\hat{\theta}_{s,k} = \theta_{s,k} + \underbrace{2\pi \frac{m_k}{N} \cdot \Delta t}_{SFO,STO} + \underbrace{\gamma}_{CFO} + Z \tag{4}$$

where $\theta_{s,k}$ is the actual phase, Δt is the time lag due to SFO and STO, m_k is the subcarrier index of the kth subcarrier, N is the discrete Fourier transform size for the OFDM generation, γ is the unknown phase offset due to CFO, and Z is the measurement noise [13].

It is worth mentioning that these offsets occur in frequency and time domains and that SFO and STO linearly depend on each subcarrier. In the following, several phase processing methods are presented to provide useful information to the models used in the field of HAR.

2.2. Linear transformation

A usual approach to mitigate offset mismatches is to apply a linear transformation. It is noticed that the phase error $2\pi \frac{m_k}{N} \Delta t + \gamma$ in Eq. (4) is a linear function of the subcarrier index m_k. We can estimate for each symbol s the phase slope ε_s and the offset τ_s with the following expressions:

$$\varepsilon_s = \frac{\hat{\theta}_{s,K} - \hat{\theta}_{s,1}}{m_K - m_1} \tag{5}$$

$$\tau_s = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_{s,k} \tag{6}$$

Finally, subtracting $\varepsilon_s m_k + \tau_s$ from the raw phase $\hat{\theta}_{s,k}$, we can obtain the calibrated phase, $\theta'_{s,k}$, which is given by

$$\theta'_{s,k} = \hat{\theta}_{s,k} - \varepsilon_s m_k - \tau_s \tag{7}$$

2.3. Savitzky-Golay filter

The Savitzky–Golay filter is a filtering method based on local area polynomial least square fitting for time-series signals [63]. It is used to smooth the CSI data and reduce environmental noise interference to facilitate the subsequent feature extraction [64]. The method requires defining a moving window of size $2l+1$ and a fitting order n to perform left-to-right curve filtering. First, the filtering center is selected, and $2l+1$ point out of each l point around the center is chosen as the primary filtering object. For the sake of simplicity, a vector $\boldsymbol{v}$ smoothed with a Savitzky–Golay filter is defined as

$$\boldsymbol{v}_{sg} = SG(\boldsymbol{v}, n, 2l+1) \tag{8}$$

where $\boldsymbol{v}_{sg}$ is the output of the filter.

The choice of the optimal parameters for the Savitzky–Golay filter depends on the nature of each problem. Therefore, they are obtained empirically through an analysis of data. In this work, the parameters of the one-dimensional SG filter have been obtained experimentally, being $n = 2$ and l is 0.1 times the length of the input vector. These are the best values for obtaining the highest accuracy for data classification (counting and activities) with this paper's evaluation methods.

Given a CSI phase matrix $\hat{\boldsymbol{\Theta}}$, SG filtering could be applied in every dimension of the matrix or in both at the same time. The CSI phase is expected to exhibit continuity across consecutive subcarriers of OFDM symbols because the channel coherence bandwidth has to be larger than the subcarrier spacing since the maximum delay spread must be much smaller than the symbol duration for the WiFi system to operate in a given environment. Furthermore, when CSI estimates of different symbols are obtained with a periodicity small enough compared with the coherence time of the wireless channel, continuity of the phase is also preserved in the inter-symbol time domain.

Therefore, this work evaluates the SG filtering in order to maintain the continuous form of the phase in the frequency domain, in the inter-symbol time domain and in both domains as follows:

1. Frequency domain: It is applied to the CSI estimate of each symbol. One of the filter characteristics is that it retains the width & height of waveform peaks in noisy signal [65].
2. Time domain: The filter is applied to each subcarrier along consecutive CSI symbols to smooth and ensure phase continuity over time. Its application has an impact on the frequency domain and can generate distortions.
3. Time–Frequency domains: Applying the filter in both domains at the same time ensures continuity and phase smoothing in both dimensions. This data processing is done according to [66].

3. Proposed method for phase processing

3.1. Phase sanitization

The idea of the proposed method is to take advantage of the good results offered by the linear transformation while maintaining continuity, i.e., avoiding gaps, in at least one of the two domains of the phase. In this sense, we improve the traditional linear transformation shown in Section 2.2 using a linear regression of overall symbol points to remove the slope generated by STO and SFO impairments. The amplitude, $\left|\hat{h}_{s,k}\right|$, remains constant and unchanged throughout the entire phase sanitation process.

Given the phase matrix $\hat{\boldsymbol{\Theta}}$, a linear regression of each s symbol (i.e., $\hat{\boldsymbol{\Theta}}_{s,*}$) in $\hat{\boldsymbol{\Theta}}$ is computed. Then, the linear regression model function follows the form:

$$r_s(k) = \varepsilon_s \cdot k + \tau_s \tag{9}$$

where τ_s is the offset:

$$\tau_s = \bar{\hat{\boldsymbol{\Theta}}}_{s,*} - \bar{k} \cdot \varepsilon_s \tag{10}$$

and ε_s is the linear regression slope:

$$\varepsilon_s = \frac{\sum_{k=1}^{K} \left(\hat{\theta}_{s,k} - \bar{\hat{\boldsymbol{\Theta}}}_{s,*}\right)(k - \bar{k})}{\sum_{k=1}^{K} (k - \bar{k})^2} \tag{11}$$

Note that $\bar{\hat{\boldsymbol{\Theta}}}_{s,*}$ and $\bar{k}$ are the average value of $\hat{\boldsymbol{\Theta}}_{s,*}$ and k, respectively.

Finally, following the shape of Eq. (7), we can obtain the corrected phased $\breve{\theta}_{s,k}$ as:

$$\breve{\theta}_{s,k} = \hat{\theta}_{s,k} - k \cdot \varepsilon_s - \tau_s \tag{12}$$

Fig. 5a shows a graphical example of the LT method and the proposed Linear Regression Transformation (LRT) solution for a specific CSI frame in the OPERAnet dataset [67].

3.2. Time smoothing and frequency rebuild

At this point, calibrated phases maintain distortions and gaps between adjacent CSI symbols in $\breve{\boldsymbol{\Theta}}$. Moreover, applying LT or LRT methods cannot ensure phase continuity in frequency, since other non-linear errors in hardware, software, or a weak implementation of the measurements can also generate gaps and deform the received signal and, in consequence, the estimated CSI. For this reason, a low-pass filter is used to smooth the calibrated CSI phases and ensure phase continuity. Time domain gaps correction makes sense if the activity to be measured generates changes in the channel at a rate greater than the time interval between OFDM symbols, which is the case in this work and generally in the HAR field.

Time Smoothing and Frequency Rebuild, TSFR, is the method proposed in this section. Assuming that the calibrated phase is approximately continuous in frequency and the main discontinuities appear in the time domain between adjacent symbols, SG filtering is proposed to be applied in the time domain, combined with a threshold-based method to correct the irregularities that SG filtering generates in the frequency domain and, thus, to maintain continuity.

Once $\breve{\boldsymbol{\Theta}}_{SxK}$ have been calculated, SG filtering is carried out in the time domain subcarrier as:

$$\breve{\boldsymbol{\Phi}}_{*,k} = SG\left(\breve{\boldsymbol{\Theta}}_{*,k},\ 2,\ 0.1 \cdot S\right) \tag{13}$$

Due to the previous time filtering, discontinuities in the frequency domain are generated in the form of a step between subcarrier blocks. A threshold-based method is proposed to remove those quantitatively large gaps that can appear between two adjacent subcarriers. In most scenarios, it can be assumed that the phase of the channel frequency response change slowly between adjacent subcarriers. As a result, the phase difference between adjacent subcarriers (i.e., $\breve{\theta}_{s,k} - \breve{\theta}_{s,k-1}$) should be small and large gaps could be considered outliers. Considering that those noisy differences can be approximated to a Gaussian distribution [56], we have defined a threshold, d_s, which has the form:

$$d_s = \mu_s + \sigma_s \tag{14}$$

where μ_s is the average of the phase differences before SG filtering:

$$\mu_s = \frac{1}{K-1} \sum_{k=2}^{K} \left|\breve{\theta}_{s,k} - \breve{\theta}_{s,k-1}\right| \tag{15}$$

and σ_s is the standard deviation:

$$\sigma_s = \sqrt{\frac{\sum_{k=2}^{K} \left(\left|\breve{\theta}_{s,k} - \breve{\theta}_{s,k-1}\right| - \mu_s\right)^2}{K-1}}. \tag{16}$$

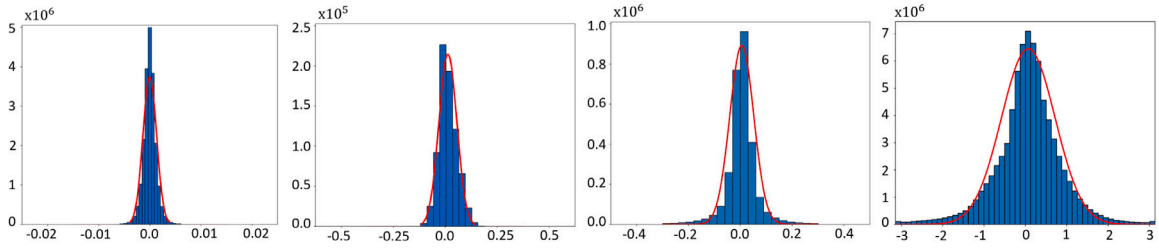


Fig. 1. Histograms and Gaussian distribution approximations of $\breve{\theta}_{s,k} - \breve{\theta}_{s,k-1}$ for HAR datasets described in Section 4.

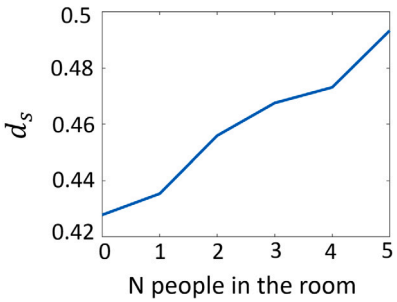


Fig. 2. Mean value of d_s for each class. In this case, it increases with the number of people in the room. These data belong to the OPERAnet dataset: tx3rx3.

Consequently, after SG filtering in the time domain, a TSFR phase matrix $\widetilde{\boldsymbol{\Phi}}_{SxK}$ is calculated where the (s,k)th element follows:

$$\widetilde{\phi}_{s,k} = \begin{cases} \breve{\phi}_{s,k} & \text{if } k = 1 \\ \widetilde{\phi}_{s,k-1} - d_s & \text{if } \epsilon < -d_s \\ \widetilde{\phi}_{s,k-1} + d_s & \text{if } \epsilon > d_s \\ \breve{\phi}_{s,k} - (\breve{\phi}_{s,k-1} - \widetilde{\phi}_{s,k-1}) & \text{otherwise} \end{cases} \tag{17}$$

being $\epsilon = \breve{\phi}_{s,k} - \breve{\phi}_{s,k-1}$.

In Fig. 1, several histograms of the phase differences of adjacent subcarriers are drawn for different datasets, which are described in Section 4. One can observe that phase difference distributions present a bell-shape and can be approximated to a Gaussian distribution.

According to this methodology, the phase of any subcarrier in which the difference with the previous one exceeds d_s will be modified. Based on the Gaussian assumption, approximately 30% of the subcarriers of each symbol are modified, including outliers generated by the SG filtering and actual smoothed values. Therefore, this methodology is not only intended to correct the outliers due to temporal filtering. It also tries to take advantage of this correction to modify the statistical distribution of the symbol, making it more characteristic for each activity by means of the d_s value. The main ideas behind this method are:

- The time evolution of the phase for each subcarrier can reveal information related to the channel variations of each activity. Therefore, those subcarriers that suffer phase gaps after time filtering are also characteristics of the time evolution of the whole CSI phase matrix, as shown in Fig. 3. One can observe that some phase differences can be sensitive to the channel changes related to the activity in the room, while others behave steady. The key point is that the threshold set makes the processed phases behave differently for each class.
- The corrected phases after the LRT method in each CSI symbol also contain relevant information related to the channel variations of each activity. Part of this information is present in its statistical variables, such as those referred to Eqs. (14), (15) and (16). The gaps generated as a result of the temporal filtering can corrupt this valuable information for HAR and, therefore, these gaps are reduced through the proposed adjustment in Eq. (17).

With this in mind, the d_s value incorporates information related to each activity into the CSI phase matrix, as is depicted in Fig. 2, generating a characteristic modal number (d_s) for each symbol and applying it in the time domain via time characteristic subcarriers k, on which condition $|\breve{\phi}_{s,k} - \breve{\phi}_{s,k-1}| > d_s$ is satisfied. With this phase correction method, the jumps are not completely eliminated, but their value is reduced and uniformed to the d_s value. So the information is preserved while distortion is reduced. In

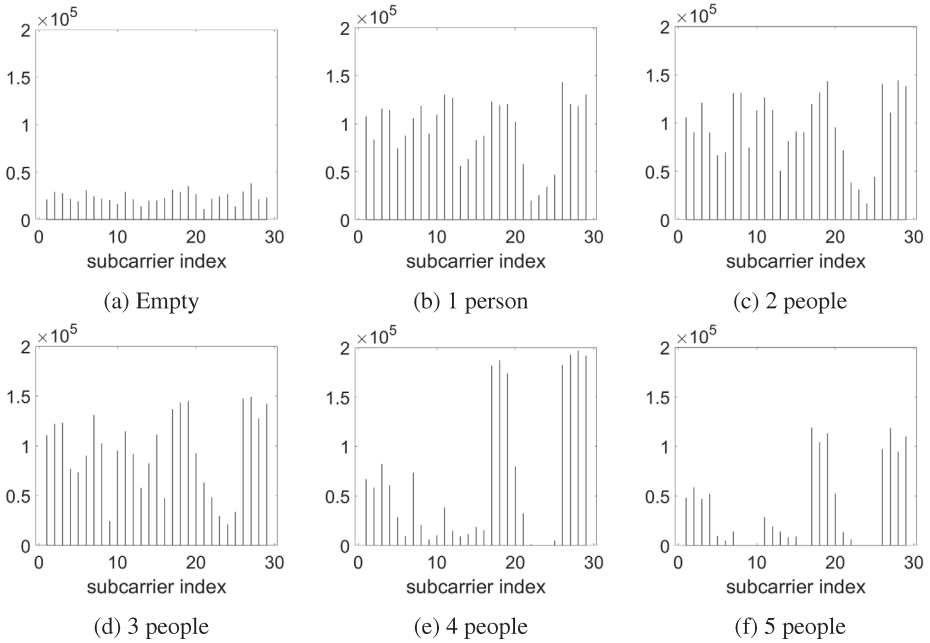


Fig. 3. Frequency distributions of the subcarriers, k, on which condition $|\breve{\phi}_{s,k} - \breve{\phi}_{s,k-1}| > d_s$ is satisfied for different number of people in a given room. These data belong to the OPERAnet dataset: tx3rx3.

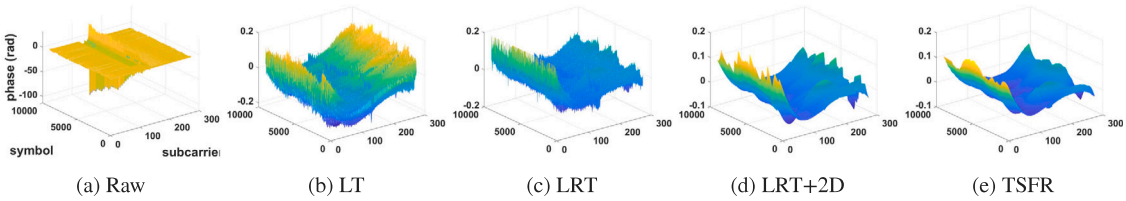


Fig. 4. Processed phase of CSI matrices with LT, LRT, LRT+2D SG filtering, and TSFR, using Dataset A.

short, new information is added to each OFDM symbol using the new variable d_s: how many times it is repeated, between which subcarriers, and what its magnitude is. All this is intended to help the prediction algorithms to classify correctly. The benefits of this processing are confirmed by the good results obtained, as seen in Section 5.

In Fig. 5b, the effect of the time SG filtering and the gap removal process is shown for a certain CSI symbol in the OPERAnet dataset (tx1rx1, $s = 50$). We can observe that several large steps are generated after SG filtering in the blue areas and, afterward, removed with the proposed threshold-based method.

The complete TSFR method is described in Algorithm 1. In Fig. 4, representations of the processed CSI phase matrices using different phase processing are shown for a certain estimated phase CSI matrix $\widehat{\boldsymbol{\Theta}}_{SxK}$ corresponding to real measurements. One can initially observe the synchronization errors in the measured phases. Corrections of the linear phase impairments are carried out with the proposed LRT solution, and $\breve{\boldsymbol{\Theta}}_{SxK}$ matrix is depicted in Fig. 4c. In Fig. 4b, we can also observe the corrections performed with the traditional LT method. Finally, the output of the TSFR solution $\widetilde{\boldsymbol{\Phi}}_{SxK}$ is given in Fig. 4e. Additionally, we can see in Fig. 4d the processed phase when the LRT method is applied along with two-dimensional SG filtering.

Finally, after the TSFR-based phase processing, the processed CSI matrix $\widetilde{\boldsymbol{H}}_{SxK}$ can be reconstructed:

$$\widetilde{\boldsymbol{H}}_{SxK} = \begin{pmatrix} \widetilde{h}_{1,1} & \cdots & \widetilde{h}_{1,K} \\ \vdots & \ddots & \vdots \\ \widetilde{h}_{S,1} & \cdots & \widetilde{h}_{S,K} \end{pmatrix} \tag{18}$$

where $\widetilde{h}_{s,k} = \left|\widehat{h}_{s,k}\right| e^{j\widetilde{\phi}_{s,k}}$.

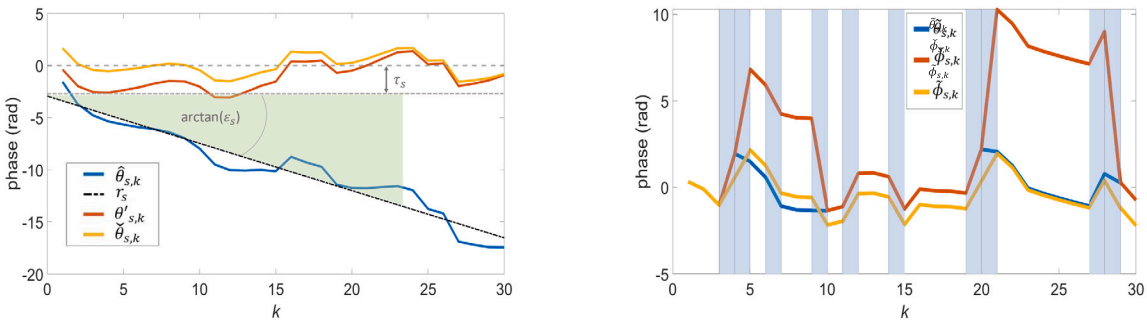


Fig. 5. (a) Graphical representation of a CFO, SFO, and STO correction in a symbol phase by LRT $\breve{\theta}_{s,k}$ vs. bibliographic method $\theta'_{s,k}$. These data belong to the OPERAnet dataset: tx1rx1, s = 500K. (b) Phases after LRT (blue), time SG smoothing (red), and TSFR (yellow). Blue areas show adjacent subcarriers where $|\breve{\phi}_{s,k} - \breve{\phi}_{s,k-1}| > d_s$ is satisfied. These data belong to the OPERAnet dataset: tx1rx1, s = 50.

Algorithm 1: TSFR

Input: $\hat{\boldsymbol{\Theta}}_{SxK}$;
Output: $\widetilde{\boldsymbol{\Phi}}_{SxK}$;
$\breve{\boldsymbol{\Theta}}_{SxK} = \boldsymbol{O}_{SxK}$;
$\breve{\boldsymbol{\Phi}}_{SxK} = \boldsymbol{O}_{SxK}$;
$\widetilde{\boldsymbol{\Phi}}_{SxK} = \boldsymbol{O}_{SxK}$;
for s *in 1:S* **do**
  $\hat{\theta}_s \leftarrow$ unwrap phase of $\hat{\boldsymbol{\Theta}}_{s,*}$;
  $b_s \leftarrow$ apply Eq. (10);
  $a_s \leftarrow$ apply Eq. (11);
  $\breve{\theta}_s \leftarrow$ apply Eq. (12);
  $\breve{\boldsymbol{\Theta}}_{s,*} = \breve{\theta}_s$;
end
for k *in 1:K* **do**
  $\breve{\theta}_k \leftarrow$ unwrap phase of $\breve{\boldsymbol{\Theta}}_{*,k}$;
  $\breve{\boldsymbol{\phi}}_k \leftarrow$ smooth $\breve{\theta}_k$ applying Eq. (13);
  $\breve{\boldsymbol{\Phi}}_{*,k} = \breve{\boldsymbol{\phi}}_k$
end
for s *in 1:S* **do**
  $\breve{\boldsymbol{\phi}}_s \leftarrow$ unwrap $\breve{\boldsymbol{\Phi}}_{s,*}$;
  $\breve{\theta}_s \leftarrow$ unwrap phase of $\breve{\boldsymbol{\Theta}}_{s,*}$
  $\mu_s \leftarrow$ apply Eq. (15);
  $\sigma_s \leftarrow$ apply Eq. (16);
  $d_s \leftarrow$ apply Eq. (14);
  $\widetilde{\boldsymbol{\phi}}_s \leftarrow$ apply Eq. (17);
  $\widetilde{\boldsymbol{\Phi}}_{s,*} = \widetilde{\boldsymbol{\phi}}_s$;
end

Table 1
Datasets.

Name	System	Class	Scenarios	RXs
Dataset A	DVB-T2 based	Counting and fixed position	1	2
Dataset B	WiFi	Counting and fixed position	1	1
EHUCount	WiFi	Counting	5	1
OPERAnet	WiFi	Counting	1	3
ReWiS	WiFi	Activities	3	1

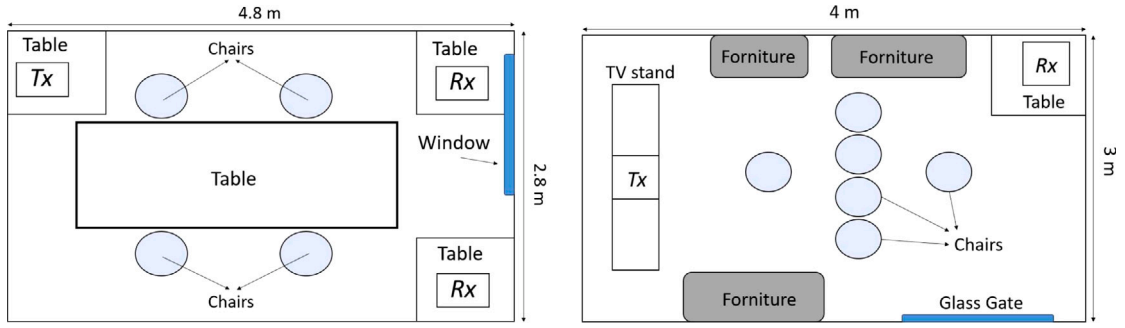
4. Evaluation setup description

4.1. Datasets

This section explains the datasets used to test the proposed phase processing method for different human activity recognition (i.e., people counting people, position detection) and in indoor environments. There are five datasets. Two of them (named A and B) are not published and are available under request. The other three are public, and their characteristics are described in detail in their respective papers [10,22,67]. The main characteristics of the datasets are shown in Table 1 and are explained below, especially for the ones that are not publicly accessible.

4.1.1. Dataset A

Dataset A is a dataset created by our research group at the University of the Basque Country. The dataset's purpose was to count people and detect their fixed positions (sitting) in an indoor environment. The measurements were taken in a meeting room



(a) Dataset A (b) Dataset B

Fig. 6. (a) Experimental room for Dataset A. Measurements have been done from empty room until 4 people, one in each chair. For each number of people, all the possible occupancy of the chairs were measured. (b) Experimental room for Dataset B. One chair is placed in each of the six selected measurement locations. Measurements were made with one or two people in the room, with one person by location, covering all the possible combinations of locations and number of people.

(2.8 m × 4.8 m) with one TX and two RXs. There were four chairs around a table in the room. The TX and the RXs were at the same height as the table. Measurements were taken in the presence of zero to four people. For each number of people, all the possible occupancy of the chairs were measured, e.g., with two people in the room, six measurements were performed, each time occupying different chairs. This makes a total of 16 measurements. The set-up for the measurements is sketched in Fig. 6A.

Three USRPs (Universal Software Radio Peripheral) were used for these measurements, one as a TX and two as RXs. A DVB-T2, 10 MHz, 32K Digital Terrestrial Television (DTT) based signal was employed for the measurements. The channel frequency was 5.4 GHz, and the sampling frequency of the TX and the RX USRPs was doubled to obtain 20 MHz bandwidth (BW). A software (SW) DVB-T2 receiver was used to decode the T2 signal and obtain the CSI, which were then decimated to work with $K = 273$ subcarriers at a rate of 606 Hz.

4.1.2. Dataset B

This dataset was created by researchers at the National Autonomous University of Mexico (UNAM). The measurements were taken in the living room of the researcher's apartment (approximately 3 × 4 m). Six different locations were selected in the room, and a chair was placed in each location. In this dataset, 37 classes are classified: one for the vacuum, one for each of the six positions that have been determined, and one for the combination of measurements with two persons, one in each position, covering all 30 possibilities. Measurements of the room without people taken only 1% of the measurements in contrast with 50% of measurements with one person or 49% of two people, so the dataset is strongly unbalanced. The room and the locations of the chairs are shown in Fig. 6B.

The measurement system consisted of two laptops with Qualcomm Atheros QCWB335 network interface cards (NIC). One of the laptops injected WiFi packets, and the other received the signal and recorded the CSI. The Atheros-CSI-Tool [68] was used for this purpose. Channel 11 of the 2.4 GHz WiFi band was used with 20 MHz BW (56 subcarriers). An average number of 50K packages were recorded, and the measurement time ranged from 13 to 18 s.

4.1.3. EHUCount

This dataset was obtained from measurements taken at the facilities of the Faculty of Engineering of the University of the Basque Country (Bilbao, Spain). The portable test bench consisted of a vector signal generator that was used to transmit a 15 s long pre-recorded IEEE 802.11n trace in the 2.4 GHz band with 20 MHz BW. The reception was performed by recording the signal as IQ samples with a signal analyzer to obtain the CSI using an SW WiFi demodulator. This dataset provides CSI from $K = 52$ subcarriers of the OFDM signal.

Measurements were carried out in five indoor scenarios where up to five people walked casually. The number of CSI traces per number of people and scenario ranged between 12K and 15K, depending on synchronization issues in the signal decoding process.

4.1.4. OPERAnet

It is a comprehensive dataset intended to evaluate passive HAR and localization techniques with measurements obtained from synchronized Radio-Frequency devices and vision-based sensors. For our purposes, the dataset consists of CSI data extracted from a WiFi NIC. Of the vast number of measurements and experiments in this dataset, we only used one, named "exp028: Crowd counting". The "exp028" dataset contains the CSI from the three TX antennas to each of the three RX antennas. For example, the CSI matrix generated between TX antenna two and RX antenna two is called tx2rx2. For convenience, only tx1rx1, tx2rx2, and tx3rx3 data have been used in this work.

For the experiment, a maximum of six people walked continuously and randomly through a room. It started with six people; then, every 5 min, one person left the monitoring area. The WiFi CSI system consisted of three PCs fitted with an Intel5300 NIC, which extracts CSI from $K = 30$ subcarriers, spread evenly among the 56 subcarriers of the 20 MHz channel 149 in the 5 GHz band at a rate of 1.6 kHz.

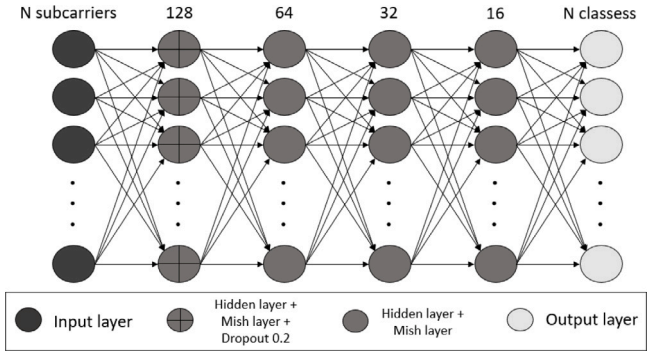


Fig. 7. Fully-connected neural network as used in this work for WiFi networks.

4.1.5. ReWiS

These measurements were carried out in three different settings. The experiments involved two subjects who were given instructions on the type, duration, and location of activities such as jumping, walking, and standing. Each measurement campaign involved 180 s of data collection for each activity performed by the two people. Measurements were repeated ten times with a time interval of at least 2 h between measurements. For the generation of the ReWiS dataset, the authors used three Asus RT-AC86U WiFi routers, each equipped with four antennas. The routers extracted the CSI packets using the Nexmon firmware [69]. The CSIs were calculated at a rate of 100 Hz, in the 5 GHz band, for 20 and 80 MHz BW, with $K = 52$ and $K = 242$ subcarriers, respectively.

4.2. Deep learning models

To test the proposed method and quantify the improvement over the bibliographic LT method described in Section 2, the datasets are manipulated in two different ways. The datasets Dataset A, Dataset B, EHUCount, and OPERAnet are evaluated by applying, on the one hand, a fully-connected neural network (FNN) and, on the other hand, a convolutional neural network (CNN). Stratified shuffle split cross-validation [70] with 5 iterations are used in the training of both networks.

In turn, the ReWiS dataset is evaluated using ProtoNet [61], a few shot learning (FSL) strategy [71], as described in his work [22].

4.2.1. Fully-connected neural network

In this case, CSI phase data is classified individually per OFDM symbol, assigning each one the label that corresponds to it. This way, if the dataset has $K \times S$ dimensions, $1 \times S$ labels are assigned. The datasets are evaluated using a full-connected neural network with four hidden layers. In addition, Mish activation layers [72] are introduced between the hidden layers to improve the information transmitted by the network using one of the new functions layer developed. Finally, a dropout layer of coefficient 0.2 is placed after the first hidden layer to avoid overfitting. The number of neurons of the first, second, third, and fourth hidden layers is 128, 64, 32, and 16, respectively, for WiFi CSI. DVB-T2 numbers are 256, 128, 64, and 32. The last layer has the same neurons as the classes to be classified. An example of this FNN is depicted in Fig. 7.

4.2.2. Convolutional neural network

To consider a sufficient time interval in which environmental changes may occur, CSI data are grouped into clusters and evaluated using a CNN. In each dataset, these groups are formed by a different number of symbols. The resizing of the data in the input network to make square inputs that can be used as images must consider the number of subcarriers, which changes for each dataset. Therefore, the dimensions of the inputs of this network are (r, r, 2), where r is equal to 128 or 256 in WiFi datasets or Dataset A, respectively, and 2 is due to that phase and amplitude are used. This way, the input obtained is comparable to a two-color square image.

This network consists of a two-channel input layer of size (r, r, 2) and three two-dimensional convolutional layers with 64, 32, and 32 neurons with three max-pooling layers between them. Behind the convolutional layers is a flattened layer to vectorize the output. Then, there are two full-connected layers, one with 32 neurons and the last one with the number of classification classes. An example of this CNN is depicted in Fig. 8.

4.2.3. Few shot learning

In the case of the ReWiS dataset, the objective is to replicate the processing performed by its authors at [22], so FSL ProtoNet processing is applied to the raw data. The goal of FSL is to generalize quickly to new tasks containing only a few samples with supervised information. ProtoNet is based on the idea that there is an environment in which points are clustered around a single prototypical representation for each class.

In this case, four activities are classified: empty, walk, stand, and jump. First, each activity set is divided into intervals of 300 symbols. Single Value Decomposition is applied to each interval to reduce its dimension from SxK to KxK. Finally, the linear correlation coefficient, or Pearson's coefficient, is applied to this matrix, obtaining another matrix of linear coefficients, KxK, which

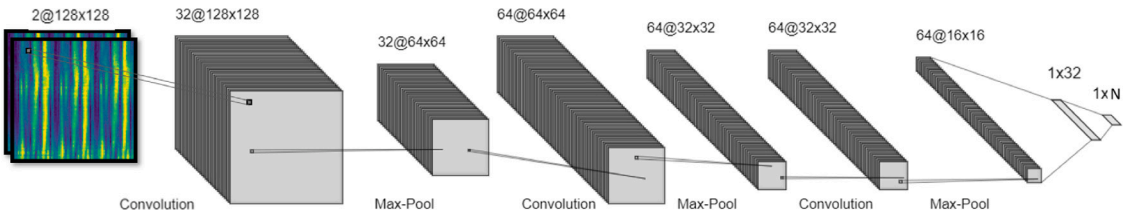


Fig. 8. Example of a Convolutional Neural Network as used in this work, with an input of size (128,128,2) and output of size N classes.

is the training network's input. The training is performed by means of a CNN with four convolutional blocks. Each block comprises a 64-filter 3×3 convolution, a batch normalization layer, a ReLU nonlinearity, and a 2×2 max-pooling layer that is applied after each of the blocks.

5. Results and discussion

In this section, several comparative analyses are carried out using the aforementioned datasets and DL models. To do that, different classification problems of HAR (people counting, position detection, and gesture recognition) are solved through the CSI information, i.e., using CSI amplitude, CSI phase, or combining amplitude and phase. Furthermore, six different phase processing methods (LT, LRT, LRT + SG filtering in the frequency domain, LRT + SG filtering in the time domain, LRT + two-dimensional SG filtering, and the TSFR proposal) are compared when CSI phases feed the described DL models. Accuracies given by the raw values of the CSI are considered as a benchmark of the models. Performance results are given in terms of averaged percentage values of accuracy along with the standard deviation. The average and the standard deviation are computed when several scenarios or receivers are provided in the same dataset.

Tables 2 and 3 show the accuracy results for people counting and position detection, respectively, using the FNN model. In this network, the values of amplitude and phase, as well as the combination of both, are used separately. The amplitude accuracy is constant in all columns of Tables 2 and 3 as it does not change after phase processing. Firstly, Table 2 shows that using amplitude versus raw phase gives better values for the Dataset A and OPERAnet datasets. In contrast, the results for Dataset B and EHUCount are similar. It is also noted that the LRT calibration of the phase gives better results than the LT method for two of the four datasets, while the accuracy is the same for the other two. Regarding smoothing, the SG filter gives better results for the time than the frequency in three of the four datasets analyzed. Still, in OPERAnet, the time smoothing generates a very low accuracy. On the other hand, 2D smoothing gives better results than frequency or time in all cases. However, the TSFR method using only the phase is the one that offers the best results of all, maintaining in all cases accuracies above 94%. Table 3 shows the results for classifying fixed positions in Datasets A and B. In this case, Dataset A consists of 17 classes, while for Dataset B it increases to 37 (Section 4.1). It can be seen that there is a big difference in the classification capacity of the FNN model in both datasets since using only the amplitude, Dataset A allows an accuracy of 80%. In comparison, Dataset B remains at 24%, similar to the other phase processing methods. The high number of classes measured in Dataset B probably limits this model's capacity. However, in Dataset B, the TSFR method demonstrates its high capacity to improve the network classification since a high increase of up to 99% is observed. In Dataset A, TSFR also offers a high accuracy of 96%.

In Tables 4 and 5, the performance of counting people and detecting position, respectively, is given for the proposed CNN model. In this case, the components of the CSI have been windowed to create images as inputs of the model. This network combines amplitude and phase values in matrices of the form $(r, r, 2)$, as explained in Section 4.2. In these tables, we can see performance indicators similar to those in Tables 2 and 3. First, the LRT method still offers better or equal accuracies than the LT method.

In Table 4, comparing time smoothing with frequency smoothing shows that one offers better results in two datasets and the other in the other two, with substantial differences. Also, in this case, 2D smoothing improves the frequency or time smoothing results in all cases, but it is outperformed by the TSFR method, which achieves excellent accuracy of more than 90% in all cases, and close to 100% in Dataset B and OPERAnet. In this table, it is striking that for Dataset B, the model classifies with an accuracy of more than 97% regardless of the phase processing used, even for the raw data. As shown in Table 6 and discussed below, this is related to the fact that there are only three classes, and in reality, it is only classifying two classes well, as the empty class only has 1% of the data.

Table 5 shows the most different results between methods concerning the other tables. For example, frequency smoothing for Dataset A gives better results than time smoothing or 2D smoothing, while for Dataset B, time smoothing is the best. For example, for Dataset A, frequency smoothing gives better results than time smoothing or 2D smoothing, while for Dataset B, time smoothing is the best. Moreover, the accuracy values in both datasets are the lowest of all.

Nevertheless, the TSFR method remains the best, with reasonable accuracy values. The accuracy of the TSFR method in Dataset B is 96%, compared to the second highest, 33%, and in Dataset A is 88%, while the second highest is 68%.

In addition, we have seen fit to include Table 6 to show the Dataset B metrics in more detail, as the high accuracy in Tables 2, and 4 can be misleading. Dataset B is an unbalanced dataset in which class *0* occupies 1% of the total size, while classes *1* and *2*

Table 2
Accuracy values, in %, of all datasets using a FNN and classifying by number of people in the room.

Dataset	Variable	Raw	LT	LRT	LRT + SG freq	LRT + SG time	LRT + SG 2D	TSFR
Dataset A	abs	74 ± 5	74 ± 5	74 ± 5	74 ± 5	74 ± 5	74 ± 5	74 ± 5
	phase	65 ± 4	81 ± 5	86 ± 1	85 ± 3	96 ± 1	95 ± 1	**97 ± 1**
	abs+phase	72 ± 8	88 ± 3	87 ± 1	89 ± 1	93 ± 5	96 ± 1	94 ± 1
Dataset B	abs	98.1	98.1	98.1	98.1	98.1	98.1	98.1
	phase	98.2	98.3	98.5	98.3	98.7	98.7	**99.9**
	abs+phase	98.4	98.3	98.2	98.1	98.7	98.3	99.3
EHUCount	abs	80 ± 7	80 ± 7	80 ± 7	80 ± 7	80 ± 7	80 ± 7	80 ± 7
	phase	79 ± 9	79 ± 9	86 ± 5	87 ± 5	**99.7 ± 0.1**	**99.8 ± 0.1**	**99.7 ± 0.1**
	abs+phase	84 ± 7	85 ± 7	88 ± 4	89 ± 2	**99.8 ± 0.2**	98.8 ± 0.8	**99.6 ± 0.2**
OPERAnet	abs	82 ± 5	82 ± 5	82 ± 5	82 ± 5	82 ± 5	82 ± 5	82 ± 5
	phase	50 ± 2	54 ± 5	54 ± 4	81 ± 3	30 ± 1	88 ± 1	**94 ± 1**
	abs+phase	65 ± 1	65 ± 1	65 ± 1	84 ± 1	64 ± 1	85 ± 1	84 ± 1

Table 3
Accuracy values, in %, of Datasets A and B using a FNN and classifying by fixed position of people in the room.

Dataset	Variable	Raw	LT	LRT	LRT + SG freq	LRT + SG time	LRT + SG 2D	TSFR
Dataset A	abs	80 ± 3	80 ± 3	80 ± 3	80 ± 3	80 ± 3	80 ± 3	80 ± 3
	phase	61 ± 1	84 ± 1	86 ± 2	86 ± 2	95 ± 1	**97 ± 1**	**96 ± 1**
	abs+phase	75 ± 5	86 ± 3	83 ± 4	90 ± 1	90 ± 3	91 ± 2	91 ± 2
Dataset B	abs	24	24	24	24	24	24	24
	phase	24	24	26	25	33	26	**99**
	abs+phase	24	23	24	26	30	25	91

Table 4
Accuracy values, in %, of all datasets using a CNN and classifying by number of people in the room.

Dataset	Raw	LT	LRT	LRT + SG freq	LRT + SG time	LRT + SG 2D	TSFR
Dataset A	46 ± 8	78 ± 5	81 ± 7	82 ± 9	53 ± 5	80 ± 7	**91 ± 3**
Dataset B	98.4	98.6	98.3	97.9	98.7	98.3	**99.9**
EHUCount	34 ± 6	55 ± 7	58 ± 4	75 ± 9	90 ± 4	94 ± 2	**96 ± 2**
OPERAnet	56 ± 3	55 ± 2	54 ± 4	55 ± 3	77 ± 1	90 ± 3	**99.9 ± 0.1**

Table 5
Accuracy values, in %, of Datasets A and B using a CNN and classifying by fixed position of people in the room.

Dataset	Raw	LT	LRT	LRT + SG freq	LRT + SG time	LRT + SG 2D	TSFR
Dataset A	29 ± 22	62 ± 17	60 ± 8	68 ± 15	42 ± 9	56 ± 6	**80 ± 8**
Dataset B	24	25	26	25	32	26	**96**

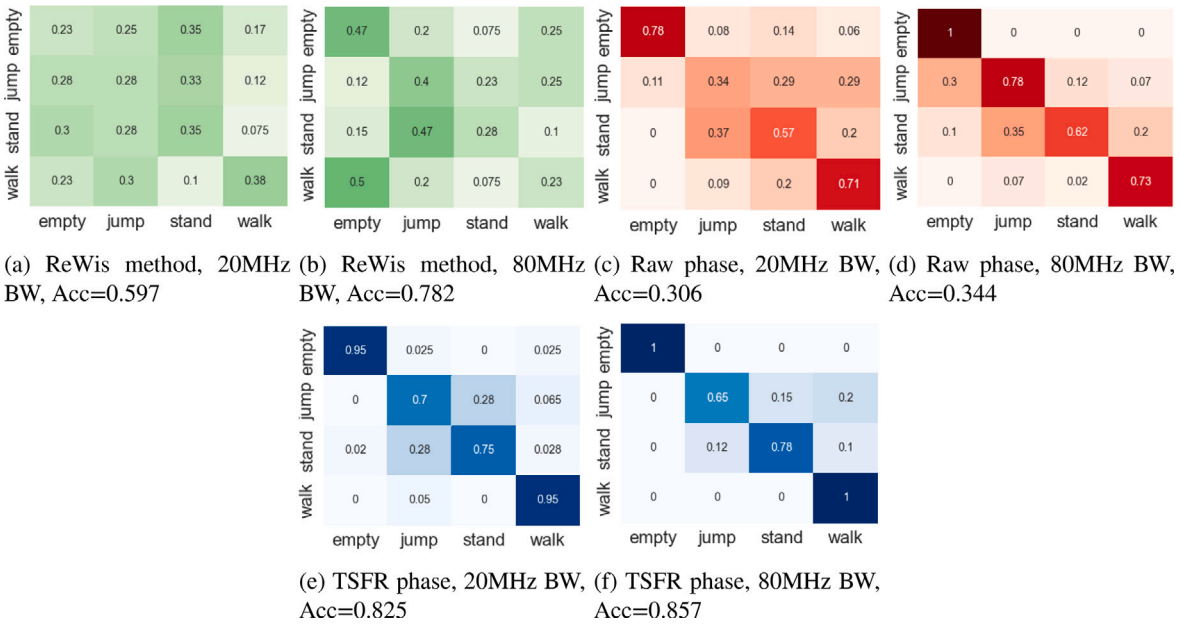
Table 6
This table is an extension of Tables 2 and 4 for Dataset B. Since this is an unbalanced dataset, accuracy results are shown for each class (0, 1, and 2 people).

Network	Variable	Raw			LT			LRT			LRT + SG freq			LRT + SG time			LRT + SG 2D			TSFR		
		0	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*
FNN	abs	0	98	98	0	99	98	0	98	99	0	98	98	0	98	98	0	98	98	0	98	98
	phase	0	98	98	0	98	98	0	98	98	0	99	98	0	98	98	0	98	99	**99**	**100**	**100**
	abs+phase	0	98	98	0	98	98	0	98	99	0	98	98	0	99	98	0	98	98	99	99	100
CNN	abs+phase	0	98	99	0	98	99	0	98	99	0	98	99	0	99	98	0	99	98	**99**	**100**	**100**

are 50% and 49%, respectively. Table 6 is an extension of Tables 2 and 4. It shows that, although the overall accuracy values are 98%, the only method capable of correctly classifying the unbalanced class is the TSFR method.

The ReWiS dataset is analyzed using FSL under the ProtoNet model. In this case, the amplitude and phase values obtained are compared using, on the one hand, the raw CSI values and, on the other hand, the CSI values processed with the TSFR method. The Fig. 9 shows the confusion matrix for each comparison for 20 and 80 MHz, including reference values from [22]. TSFR phase processing improves the CSI raw results from 32% to 82% and from 35% to 85% at 20 and 80 MHz of bandwidth, respectively, for phase accuracy. These results outperform the accuracy using the amplitude. Testing the TSFR method on this dataset using FSL implies that the method supports the extraction of certain features on the processed phase and improves the transferability of its results between different scenarios.

To summarize, the results indicate that the TSFR method can improve the classification accuracy by counting people, determining their fixed position, and detecting activities using regular neural networks, as shown in all datasets. The success of the Eq. (17)



(a) ReWis method, 20MHz BW, Acc=0.597 (b) ReWis method, 80MHz BW, Acc=0.782 (c) Raw phase, 20MHz BW, Acc=0.306 (d) Raw phase, 80MHz BW, Acc=0.344

(e) TSFR phase, 20MHz BW, Acc=0.825 (f) TSFR phase, 80MHz BW, Acc=0.857

Fig. 9. Confusion matrices: (a) and (b) are based on the ReWis method [22] using CSI amplitude. (c) and (d) are achieved with the raw CSI phase. (e) and (f) are based on the TSFR method. The data corresponds to the configuration of single antennas in transmission and reception.

is observed by comparing the results of the SG filter in the time domain (LRT+SG time) vs. the TSFR method. In all cases, the proposed method in Eq. (17) to rebuild the distortions generated by the SG filtering in the frequency domain substantially improves the classification algorithms. The accuracy of the TSFR method is always higher than that given by the exclusive application of the SG filter in any domain, including both simultaneously. Moreover, TSFR obtains good results when DL algorithms use CSI data directly, and it also improves results when feature engineering is carried out, as we have observed in the comparative analysis based on the FSL model. Furthermore, it is worth mentioning that the results seem to indicate that their use in unbalanced datasets may help to improve the accuracy in detecting under-represented classes.

6. Conclusion & future work

This paper introduces the CSI phase processing method called Time Smoothing and Frequency Rebuild (TSFR) for wireless sensing HAR. TSFR corrects phase errors through linear regression, applies time-domain smoothing with the Savitzky–Golay filter, and rebuilds frequency distortions from sanitized phase data. The method is generalized for five datasets with different properties, varying subcarrier numbers and hardware, while classifications have been done using three DL models, including FSL. Future work can explore parameter adjustments for specific datasets, including values for the Savitzky–Golay and Gaussian filters to handle outliers and adapt to dataset specific characteristics.

In summary, this work proves that TSFR outperforms other phase processing methods across five CSI datasets from diverse wireless systems and HAR activities. The proposal has been validated with three DL networks, providing high accuracy in all cases and confirming its robustness for HAR in varied wireless sensing conditions.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Guillermo Diaz reports financial support was provided by Basque Government.

Data availability

Some data are available under request, other is public and published.

Funding

This work has been financially supported by the Basque Government (under grant IT1436-22) and by the Spanish Government (under grant PID2021-124706OB-I00, funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe).

References

[1] R. Saini, P. Kumar, B. Kaur, P.P. Roy, D.P. Dogra, K.C. Santosh, Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare, Int. J. Mach. Learn. Cybern. 10 (2019) 2529–2540.

[2] S. Bei, Z. Zhen, Z. Xing, L. Taocheng, L. Qin, Movement disorder detection via adaptively fused gait analysis based on kinect sensors, IEEE Sens. J. 18 (17) (2018) 7305–7314.

[3] G. Díaz-SanMartín, L. Reyes-González, S. Sainz-Ruiz, L. Rodríguez-Cobo, J.M. López-Higuera, Automatic ankle angle detection by integrated RGB and depth camera system, Sensors 21 (2021) 1–21.

[4] I. Alrashdi, M.H. Siddiqi, Y. Alhwaiti, M. Alruwaili, M. Azad, Maximum entropy Markov model for human activity recognition using depth camera, IEEE Access 9 (2021) 160635–160645.

[5] W. Wei, F. Tan, H. Zhang, H. Mao, M. Fu, O.W. Samuel, G. Li, Surface electromyogram, kinematic, and kinetic dataset of lower limb walking for movement intent recognition, Sci. Data 10 (2023).

[6] M.Z. Uddin, A. Soylu, Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning, Sci. Rep. 11 (2021).

[7] P. Kumar, S. Suresh, DeepTransHAR: a novel clustering-based transfer learning approach for recognizing the cross-domain human activities using GRUs (gated recurrent units) networks, Internet Things 21 (2023) 100681.

[8] S. Hosseininoorbin, S. Layeghy, B. Kusy, R. Jurdak, M. Portmann, HARBIC: Human activity recognition using bi-stream convolutional neural network with dual joint time–frequency representation, Internet Things 22 (2023) 100816.

[9] L. Yao, Q.Z. Sheng, X. Li, T. Gu, M. Tan, X. Wang, S. Wang, W. Ruan, Compressive representation for device-free activity recognition with passive RFID signal strength, IEEE Trans. Mob. Comput. 17 (2) (2018) 293–306.

[10] I. Sobron, J.D. Ser, I. Eizmendi, M. Velez, Device-free people counting in IoT environments: New insights, results, and open challenges, IEEE Internet Things J. 5 (2018) 4396–4408.

[11] J.R.M. Bernaola, I. Sobrón, J. Del Ser, I. Landa, I. Eizmendi, M. Vélez, Ensemble learning for seated people counting using WiFi signals: Performance study and transferability assessment, in: 2021 IEEE Globecom Workshops (GC Wkshps), 2021, pp. 1–6, http://dx.doi.org/10.1109/GCWkshps52748.2021.9682014.

[12] L. Kong, Z. Wu, G. Chen, M. Qiu, S. Mumtaz, J.J.P.C. Rodrigues, Crowdsensing-based cross-operator switch in rail transit systems, IEEE Trans. Commun. 68 (12) (2020) 7938–7947.

[13] X. Wang, L. Gao, S. Mao, CSI phase fingerprinting for indoor localization with a deep learning approach, IEEE Internet Things J. 3 (2016) 1113–1123.

[14] K. Qian, C. Wu, Z. Yang, Y. Liu, H.E. Fugui, T. Xing, Enabling contactless detection of moving humans with dynamic speeds using CSI, ACM Trans. Embed. Comput. Syst. 17 (2018).

[15] Y. Xu, W. Yang, J. Wang, X. Zhou, H. Li, L. Huang, Wistep: Device-free step counting with WiFi signals, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (4) (2018).

[16] J. Liu, L. Wang, L. Guo, J. Fang, B. Lu, W. Zhou, A research on CSI-based human motion detection in complex scenarios, in: 2017 IEEE 19th International Conference on E-Health Networking, Applications and Services (Healthcom), 2017, pp. 1–6, http://dx.doi.org/10.1109/HealthCom.2017.8210800.

[17] Y. Zeng, D. Wu, R. Gao, T. Gu, D. Zhang, FullBreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (2018) 1–19.

[18] Y. Zeng, Z. Liu, D. Wu, J. Liu, J. Zhang, D. Zhang, A multi-person respiration monitoring system using COTS wifi devices, in: UbiComp/ISWC 2020 Adjunct - Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, Association for Computing Machinery, 2020, pp. 195–198.

[19] A. Khamis, B. Kusy, C.T. Chou, W. Hu, Wirelax: Towards real-time respiratory biofeedback during meditation using WiFi, Ad Hoc Netw. 107 (2020).

[20] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, S. Li, RT-fall: A real-time and contactless fall detection system with commodity WiFi devices, IEEE Trans. Mob. Comput. 16 (2017) 511–526.

[21] S. Palipana, D. Rojas, P. Agrawal, D. Pesch, Falldefi, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (2018) 1–25.

[22] N. Bahadori, J. Ashdown, F. Restuccia, Rewis: Reliable wi-fi sensing through few-shot multi-antenna multi-receiver CSI learning, in: 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2022, pp. 50–59, http://dx.doi.org/10.1109/WoWMoM54355.2022.00027.

[23] S. Palipana, P. Agrawal, D. Pesch, Channel state information based human presence detection using non-linear techniques, in: Proceedings of the 3rd ACM Conference on Systems for Energy-Efficient Built Environments, BuildSys 2016, 2016, pp. 177–186.

[24] J. Wang, L. Zhang, Q. Gao, M. Pan, H. Wang, Device-free wireless sensing in complex scenarios using spatial structural information, IEEE Trans. Wireless Commun. 17 (2018) 2432–2442.

[25] L. Guo, L. Wang, J. Liu, W. Zhou, B. Lu, Huac: Human activity recognition using crowdsourced WiFi signals and skeleton data, Wirel. Commun. Mob. Comput. 2018 (2018).

[26] W. Wang, A.X. Liu, M. Shahzad, K. Ling, S. Lu, Device-free human activity recognition using commercial WiFi devices, IEEE J. Sel. Areas Commun. 35 (2017) 1118–1131.

[27] F. Xiao, J. Chen, X. Xie, L. Gui, L. Sun, R. Wang, SEARE: A system for exercise activity recognition and quality evaluation based on green sensing, IEEE Trans. Emerg. Top. Comput. 8 (2020) 752–761.

[28] Q. Gao, J. Wang, X. Ma, X. Feng, H. Wang, CSI-based device-free wireless localization and activity recognition using radio image features, IEEE Trans. Veh. Technol. 66 (2017) 10346–10356.

[29] I. Landa, G. Díaz, I. Sobrón, I. Eizmendi, M. Vélez, WIP: Impulsive noise source recognition with OFDM-WiFi signals based on channel state information using machine learning, in: 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2022, pp. 157–160, http://dx.doi.org/10.1109/WoWMoM54355.2022.00047.

[30] Task group bf (wlan sensing), IEEE P802.11 - task group bf (wlan sensing), 2021, URL: https://www.ieee802.org/11/Reports/tgbf_update.htm. Accessed: 2023-05-30.

[31] H. Shi, Y. Lu, J. Du, W. Du, X. Ye, X. Yu, J. Ma, J. Cheng, Y. Gao, Y. Cao, L. Zhou, Q. Li, Application of back propagation artificial neural network on genetic variants in adiponectin ADIPOQ, peroxisome proliferator-activated receptor-y, and retinoid x receptor-a genes and type 2 diabetes risk in a Chinese han population, Diabetes Technol. Ther. 14 (3) (2012) 293–300, PMID: 22023374.

[32] A.A. El Naby, E. El-Din Hemdan, A. El-Sayed, Deep learning approach for credit card fraud detection, in: 2021 International Conference on Electronic Engineering (ICEEM), 2021, pp. 1–5, http://dx.doi.org/10.1109/ICEEM52022.2021.9480639.

[33] A. Luckow, K. Kennedy, M. Ziolkowski, E. Djerekarov, M. Cook, E. Duffy, M. Schleiss, B. Vorster, E. Weill, A. Kulshrestha, M.C. Smith, Artificial intelligence and deep learning applications for automotive manufacturing, in: 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 3144–3152, http://dx.doi.org/10.1109/BigData.2018.8622357.

[34] J. Lund, Y.-K. Ng, Movie recommendations using the deep learning approach, in: 2018 IEEE International Conference on Information Reuse and Integration (IRI), 2018, pp. 47–54, http://dx.doi.org/10.1109/IRI.2018.00015.

[35] S. Jha, M. Fattah, M. Karthick, Convolutional neural networks for breast cancer detection using regions of interest from infrared images, Tamjeed J. Healthc. Eng. Sci. Technol. 1 (2) (2023) 44–53.

[36] A. Faris, M. Badamasi, Feasibility of breast cancer detection through a convolutional neural network in mammographs, Tamjeed J. Healthc. Eng. Sci. Technol. 1 (2) (2023) 36–43.

[37] D. Avola, L. Cinque, A. Diko, A. Fagioli, G.L. Foresti, A. Mecca, D. Pannone, C. Piciarelli, MS-faster R-CNN: Multi-stream backbone for improved faster R-CNN object detection and aerial tracking from UAV images, Remote Sens. 13 (9) (2021) 1670.

[38] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, D. Li, Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment, IEEE Trans. Ind. Inform. 14 (9) (2018) 4224–4231.

[39] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016).

[40] Y. Alqahtani, M. Nawaz, T. Nazir, A. Javed, F. Jeribi, A. Tahir, An improved deep learning approach for localization and recognition of plant leaf diseases, Expert Syst. Appl. 230 (2023) 120717.

[41] N. Efthymiou, P.P. Filntisis, P. Koutras, A. Tsiami, J. Hadfield, G. Potamianos, P. Maragos, Childbot: Multi-robot perception and interaction with children, Robot. Auton. Syst. 150 (2022) 103975.

[42] C. Yi, H. Chen, X. Hu, Y. Xu, Domain adaptation from public dataset to robotic perception based on deep neural network, in: 2020 Chinese Automation Congress (CAC), 2020, pp. 6218–6222, http://dx.doi.org/10.1109/CAC51589.2020.9327112.

[43] F. Tarik, K. M., Artificial intelligence innovation and human resource recruitment, Tamjeed J. Healthc. Eng. Sci. Technol. 1 (2) (2023) 20–29.

[44] J. Huang, B. Liu, H. Jin, Z. Liu, Wianti: an anti-interference activity recognition system based on WiFi CSI, in: 2018 IEEE International Conference on Internet of Things (IThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2018, pp. 58–65, http://dx.doi.org/10.1109/Cybermatics_2018.2018.00044.

[45] X. Wang, Z. Zhang, D. He, K. Guan, D. Liu, J. Dou, S. Mumtaz, S. Al-Rubaye, A multi - task learning model for super resolution of wireless channel characteristics, in: GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 2022, pp. 952–957, http://dx.doi.org/10.1109/GLOBECOM48099.2022.10001700.

[46] J. Huang, B. Liu, C. Chen, H. Jin, Z. Liu, C. Zhang, N. Yu, Towards anti-interference human activity recognition based on WiFi subcarrier correlation selection, IEEE Trans. Veh. Technol. 69 (6) (2020) 6739–6754.

[47] S. Sen, B. Radunovic, R.R. Choudhury, T. Minka, You are facing the mona lisa: Spot localization using PHY layer information, in: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys '12, 2012, pp. 183–196, http://dx.doi.org/10.1145/2307636.2307654.

[48] K. Qian, C. Wu, Z. Yang, Y. Liu, Z. Zhou, PADS: Passive detection of moving targets with dynamic speed using phy layer information, in: 2014 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS), 2014, pp. 1–8, http://dx.doi.org/10.1109/PADSW.2014.7097784.

[49] X. Wang, L. Gao, S. Mao, Phasefi: Phase fingerprinting for indoor localization with a deep learning approach, in: 2015 IEEE Global Communications Conference (GLOBECOM), 2015, pp. 1–6, http://dx.doi.org/10.1109/GLOCOM.2015.7417517.

[50] S.-H. Fang, C.-C. Li, W.-C. Lu, Z. Xu, Y.-R. Chien, Enhanced device-free human detection: Efficient learning from phase and amplitude of channel state information, IEEE Trans. Veh. Technol. 68 (3) (2019) 3048–3051.

[51] X. Dang, J. Ren, Z. Hao, Y. Hei, X. Tang, Y. Yan, A novel indoor localization method using passive phase difference fingerprinting based on channel state information, Int. J. Distrib. Sens. Netw. 15 (2019).

[52] X. Cheng, B. Huang, J. Zong, Device-free human activity recognition based on GMM-HMM using channel state information, IEEE Access 9 (2021) 76592–76601.

[53] Q. Bu, X. Ming, J. Hu, T. Zhang, J. Feng, J. Zhang, TransferSense: towards environment independent and one-shot wifi sensing, Pers. Ubiquitous Comput. 26 (2022) 555–573.

[54] M. Kotaru, K. Joshi, D. Bharadia, S. Katti, Spotfi: Decimeter level localization using WiFi, SIGCOMM Comput. Commun. Rev. 45 (4) (2015) 269–282.

[55] H. Zhu, Y. Zhuo, Q. Liu, S. Chang, π-Splicer: Perceiving accurate CSI phases with commodity WiFi devices, IEEE Trans. Mob. Comput. 17 (9) (2018) 2155–2165.

[56] N. Tadayon, M.T. Rahman, S. Han, S. Valaee, W. Yu, Decimeter ranging with channel state information, IEEE Trans. Wireless Commun. 18 (7) (2019) 3453–3468.

[57] F. Meneghello, D. Garlisi, N.D. Fabbro, I. Tinnirello, M. Rossi, SHARP: Environment and person independent activity recognition with commodity IEEE 802.11 access points, IEEE Trans. Mob. Comput. (2022) 1–16.

[58] Y. Yang, J. Cao, X. Liu, X. Liu, Wi-count: Passing people counting with COTS WiFi devices, in: 2018 27th International Conference on Computer Communication and Networks (ICCCN), 2018, pp. 1–9, http://dx.doi.org/10.1109/ICCCN.2018.8487420.

[59] H. Yang, Z. Ji, J. Sun, F. Xing, Y. Shen, W. Zhuang, W. Zhang, Recognition for human gestures based on convolutional neural network using the off-the-shelf wi-fi routers, Wirel. Commun. Mob. Comput. 17 (2021).

[60] G. Díaz, I. Sobrón, I. Eizmendi, I. Landa, M. Vélez, Channel phase calibration based on savitzky-golay filter in time-domain for OFDM systems, in: 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2022, pp. 1–4, http://dx.doi.org/10.1109/BMSB55706.2022.9828560.

[61] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Comput. Surv. 53 (2020).

[62] M. Speth, S. Fechtel, G. Fock, H. Meyr, Optimum receiver design for wireless broad-band systems using OFDM, IEEE Trans. Commun. 47 (11) (1999) 1668–1677.

[63] R.W. Schafer, What is a Savitzky-Golay filter? IEEE Signal Process. Mag. 28 (2011) 111–117.

[64] L. Jiacheng, Z. Yihong, C. Yunfang, Z. Wei, How to improve the signal processing of WiFi sensing, in: 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), 2017, pp. 912–915, http://dx.doi.org/10.1109/ICCSN.2017.8230243.

[65] S. Agarwal, A. Rani, V. Singh, Performance evaluation and implementation of FPGA based SGSF in smart diagnostic applications, J. Med. Syst. 40 (2015).

[66] S.Y. Huang, Savitzky-golay smoothing filter for 2D data, 2022, https://www.mathworks.com/matlabcentral/fileexchange/37147-savitzky-golay-smoothing-filter-for-2d-data. Accessed: 2022-08-01.

[67] M.J. Bocus, W. Li, S. Vishwakarma, R. Kou, C. Tang, K. Woodbridge, I. Craddock, R. McConville, R. Santos-Rodriguez, K. Chetty, R. Piechocki, Operanet: A multimodal activity recognition dataset acquired from radio frequency and vision-based sensors, Sci. Data (2022).

[68] Y. Xie, Z. Li, M. Li, Precise power delay profiling with commodity WiFi, in: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom '15, 2015, pp. 53–64, http://dx.doi.org/10.1145/2789168.2790124, URL: http://doi.acm.org/10.1145/2789168.2790124.

[69] M. Schulz, D. Wegemer, M. Hollick, Nexmon: The C-based firmware patching framework, 2020, https://nexmon.org. Accessed: 2022-08-22.

[70] Stratified shuffle split, stratified shuffle split, 2017, https://scikit-learn.org/stable/modules/generated/sklearn.model-selection.StratifiedShuffleSplit.html. Accessed: 2022-08-10.

[71] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS '17, 2017, pp. 4080–4090.

[72] D. Misra, Mish: A self regularized non-monotonic neural activation function, in: British Machine Vision Conference (BMVC), 2020.