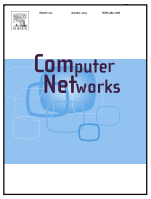
Computer Networks 275 (2026) 111907

Contents lists available at ScienceDirect

Computer Networks

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



Si-FI: Learning the beamforming feedback for simultaneous multi-subject sensing

Khandaker Foysal Haque, Milin Zhang, Francesca Meneghello, Francesco Restuccia

Institute for the Wireless Internet of Things, Northeastern University, Boston, MA, USA

ARTICLE INFO

Keywords:
Wi-Fi sensing
Simultaneous multi-subject sensing
Simultaneous wifi sensing
Beamforming feedback based wifi sensing
Simultaneous multi-subject wifi sensing

ABSTRACT

There has been significant progress in Wi-Fi sensing applications pertaining to home surveillance, remote healthcare, and home entertainment among others. However, most of the work leverages manual extraction of Channel State Information (CSI) from Wi-Fi network interface card (NIC) and targets single-subject sensing. In this work, we devise a *simultaneous multi-subject* sensing strategy that can adapt to different environments and people being monitored. `Si-FI` leverages standard-compliant beamforming feedback information (BFI) as a proxy of CSI to characterize the propagation environment. Unlike CSI, BFI (i) can be captured without any firmware modifications and (ii) captures the multiple channels between the access point and the stations without any direct access to the sensing devices. Thus, conversely, from existing work, the edge server in `Si-FI` records the BFI of the channels between Access Point (AP) and all the stations (STAs) (sensing devices) with a single capture, reducing the channel occupation, transmission and system latency dramatically. To achieve generalization over unseen environments and people, we develop a few-shot learning algorithm named `Si-FI FREL` to operate with beamforming feedback angles (BFAs) (compressed BFI). We validate `Si-FI` through an extensive data collection campaign in 3 different environments and 3 subjects performing 20 different activities simultaneously. We demonstrate that `Si-FI` achieves classification accuracy of up to 99 %, while `Si-FI FREL` improves the accuracy up to 27 % when compared to the state-of-the-art domain adaptation algorithm. `Si-FI` reduces the system latency by 50 % and channel occupation by 110 KB per sample for each sensing device compared to the state-of-the-art simultaneous multi-subject sensing work.

1. Introduction

In today's world, Wi-Fi is the technology of choice for indoor Internet connectivity. Projections suggest that Wi-Fi economic value will reach a value of $4.9 trillion by 2025. Given its widespread deployment, Wi-Fi is being investigated for applications beyond pervasive indoor connectivity. Wi-Fi devices can be used to implement a broad range of device-free sensing applications including indoor localization [1], activity recognition [2,3], and health monitoring [4]. The intuition behind Wi-Fi sensing is that the transmitted Wi-Fi signals undergo reflections, diffractions, and scattering when entities are present between the transmitter and the receiver [5–7].

Based on the high potential of this technology, Wi-Fi sensing has been identified as one of the applications that should be supported by next-generation IEEE 802.11 standards. In this view, the IEEE 802.11bf working group has been established to define the requirements for supporting sensing applications through Wi-Fi devices. Despite this effort, Wi-Fi sensing is not yet a mature technology. Current approaches are limited by firmware-specific hardware and system data overhead that limits their widespread implementation (see Section 2). Moreover, the algorithms proposed in the literature – discussed in Section 3 – do not effectively address the challenge of simultaneous multi-subject sensing, which is a key feature for smart home applications.

In this work, we present `Si-FI`, a completely novel approach for human activity recognition that is implemented on any Wi-Fi device without introducing overhead for data transmission. This is achieved by changing the sensing paradigm, moving from a CSI-based approach to a BFAs-based approach, as explained next. On top of this, in contrast to existing work, `Si-FI` supports simultaneous multiple-subject activity recognition. In `Si-FI` we address the firmware-dependency and data overhead challenges by leveraging the BFAs data exchanged during multiple-input multiple-output (MIMO) channel sounding as a sensing primitive, as a proxy to CSI data. Hence, inspired by SiMWiSense [8], we implement simultaneous multiple-subject sensing by leveraging multiple multi-user (MIMO) (MU-MIMO) STAs where the BFAs from the STA closest to each subject is utilized for activity classification of that

E-mail addresses: haque.k@northeastern.edu (K.F. Haque), zhang.mil@northeastern.edu (M. Zhang), fr.meneghello@northeastern.edu (F. Meneghello), frestuc@northeastern.edu (F. Restuccia).

https://doi.org/10.1016/j.comnet.2025.111907
Received 6 March 2025; Received in revised form 11 November 2025; Accepted 30 November 2025
Available online 3 December 2025

subject. Note that finding the closest device to a given subject falls under the Wi-Fi indoor localization and/or fingerprinting problem [9–11], which has been thoroughly investigated in the literature.

Fig. 1(a) presents a high-level comparison between Si-FI and conventional CSI-based sensing approaches. Unlike CSI-based methods, Si-FI operates without requiring direct access to the sensing devices, since BFAs-based sensing is firmware-independent and does not involve any firmware modification [12,13]. In Si-FI, the edge server passively captures the BFAs frames transmitted by the STAs associated with the subjects engaged in simultaneous sensing. Therefore, the STAs are not required to transmit any additional data to the edge server, departing from CSI-based simultaneous sensing schemes that explicitly require the STAs to offload CSI data to the edge server. As detailed in Section IV-A, according to the IEEE 802.11 standard, each STA transmits its beamforming BFAs to the AP during regular channel sounding initiated by the AP. Si-FI simply leverages these standard feedback frames without introducing any sensing overhead to transmit additional packets for sensing purposes. This design significantly reduces both system latency and channel occupation compared to CSI-based approaches.

Si-FI activity recognition algorithm is based on a Feature Reusable Embedding Learning (FREL) architecture which can adapt to any new environment or subject with only 15 seconds of new data for each sensing task. This adaptability feature is essential for plug-and-play deployments in real-world environments.

Summary of Novel Contributions

• We present Si-FI, the first BFAs-based system for simultaneous multi-subject activity classification. Unlike existing approaches, Si-FI leverages BFAs extracted from the unencrypted frames transmitted by the STAs to the AP. Capturing such frames does not require specialized tools or hardware-specific firmware. This makes Si-FI more scalable than existing CSI-based strategies. Moreover, Si-FI edge server records the BFAs from all the participating STAs with a single capture thus significantly improving the system latency and channel occupation.

• To achieve generalization to new untrained environments and subjects, we propose a novel BFAs-based Few-Shot Learning (FSL) architecture called Si-FI-FREL. In stark contrast to existing approaches, Si-FI-FREL can adapt to any new environment and subject through two main steps, namely meta-learning and fine-tuning. In contrast to the traditional FSL, FREL combines both the embedding learning and meta-learning approaches to achieve better performance through fine-tuning the classifier with only a few additional samples (see Section 4).

• We performed an exhaustive data collection campaign in 3 different environments, with 3 subjects performing 20 different activities simultaneously. We also set up another co-located network for synchronous CSI capture for comparative analysis. Si-FI achieves classification accuracy of up to 99 %, while Si-FI FREL improves the accuracy up to 27 % when compared to the state-of-the-art domain adaptation algorithms. While the activity classification performance is comparable with the state-of-the-art simultaneous activity classification approach – SiMWiSense [8] – Si-FI improves the system latency by 50 % and channel occupation by 110 KB per sample for each sensing device. **For reproducibility, we share our dataset, and the code repository at https://github.com/kfoysalhaque/Si-FI.**

The rest of the article is organized as follows. In Section 2 we discuss the key challenges of simultaneous multi-subject sensing and existing literatures are illustrated in Section 3. The Si-FI simultaneous multi-subject sensing is presented in Section 4 whereas the experimental setup and performance analysis are presented in Section 5 and Section 6 respectively. Finally Section 7 concludes the discussion.

2. Key challenges and novelties

Si-FI addresses three main key challenges that are present in current research work: scalability, firmware-dependent software, and data overhead as detailed next. Please note that “station (STA)” and “sensing device” are used interchangeably throughout the rest of the paper, both referring to the Wi-Fi nodes participating in the sensing process.

2.1. Scalability challenge

Scalability is a key challenge in simultaneous sensing. Specifically, when dealing with classification tasks – e.g., human activity classification – the system complexity grows with the number of classes to be considered. To classify m activities performed by n subjects, the total number of classes to distinguish is m^n. It is worth noting that the expression m^n denotes the total number of possible multi-subject activity combinations in the simultaneous sensing scenario, where n subjects simultaneously perform different activities. As each subject independently engages in one of the m activities, the system must distinguish among all possible combinations, leading to m^n total classes. For example, a system to identify n =10 activities for m =3 subjects would require the algorithm to learn features for more than 59,000 classes. In our recent works – SiMWiSense [8] – we addressed the challenge of simultaneous sensing by obtaining the activity estimate based only on the sensing device positioned closest to the subject. This approach relies on the intuition that the best accuracy is experienced when the transceiver positioned closest to the subject is used as a sensing device.

In Si-FI we develop a similar approach for addressing the scalability challenge, as detailed in Section 4.

2.2. Firmware-specific software challenge

This issue is specific to systems that leverage CSI data as sensing primitive, i.e., most of the systems in the literature, including SiMWiSense. CSI is computed at the physical layer (PHY) and not made available to the users at the application layer by off-the-shelf NICs. Although CSI can be extracted with software defined radio (SDR) implementations, they are costly specialized hardware that may be unavailable in real-life situations and require expert knowledge to be used. To overcome such limitations, in recent years, researchers have developed a few CSI extraction tools that run on commercial Wi-Fi NICs. Two of them, namely Linux CSI [14] and Atheros CSI [15], target IEEE 802.11n-compliant NICs (working with up to 40 MHz of bandwidth). The newer Nexmon CSI [16] allows extracting the channel frequency response (CFR)[1] from a few IEEE 802.11ac compliant devices, supporting bandwidths up to 80 MHz. The most recent one, AX CSI [17] is designed for IEEE 802.11ax devices and allows obtaining CFR measurements also on 160 MHz bandwidth channels. These tools, however, need non-trivial firmware modifications of the NICs which limits their widespread adoption for Wi-Fi sensing. Moreover, they do not provide support for estimating the CSI on MU-MIMO channels – they only monitor the MIMO links between the transmitter and the device implementing the tool, i.e., only single-user (MIMO) (SU-MIMO) mode is supported.

As shown in Fig. 1, Si-FI addresses this challenge by using the BFAs as sensing primitive instead of relying on the CSI. While CSI extraction tools are firmware-specific, BFAs extractors, e.g., Wi-BFI [13] can be implemented on any Wi-Fi-compliant devices.

2.3. Data overhead challenge

The CSI tools [14–17] capture and store the CSI representing the MIMO channel between the device performing the estimation and the transmitter. In most cases, the collected data needs to be transmitted from each device to the edge server for sensing task execution as deep learning (DL) techniques or advanced signal processing methods are usually adopted and sensing devices (such as AP, smartphones) are battery and memory constrained. This causes over-the-air overhead for

[1] CSI and CFR are used interchangeability in this article.

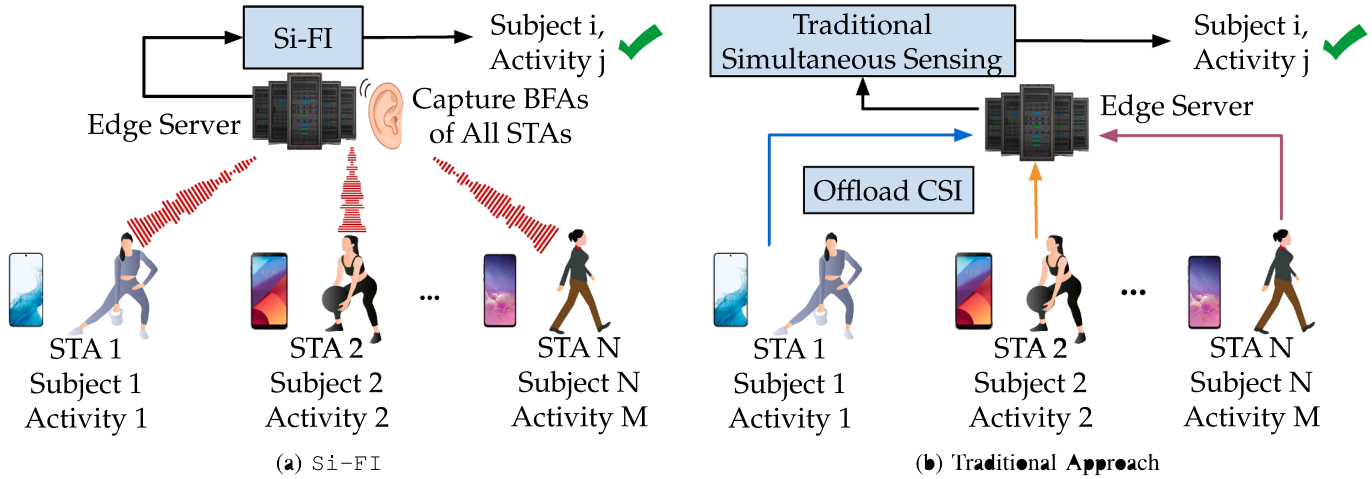


Fig. 1. Si-FI vs traditional simultaneous sensing.

sensing data transmission that increases channel occupation and hampers the sensing performance by increasing the overall system latency. Moreover, for simultaneous multi-subject sensing, all the CSI collecting devices within the sensing system need to be time-synchronized increasing system complexity. By relying on the BFAs instead of the CSI, Si-FI effectively addresses this challenge.

2.4. Main novelties of Si-FI

Si-FI brings transformative advancements in Wi-Fi-based simultaneous multi-subject activity classification, building upon its predecessor – SiMWiSense [8] and addressing critical challenges in scalability, adaptability, computational efficiency, and data overhead as presented in Section 6. One of the standout features of Si-FI is its ability to manage the exponential growth of classes inherent in multi-subject activity sensing. By assigning each subject to the closest sensing device, Si-FI ensures minimal channel interference and significantly enhances classification accuracy. This decentralized approach has been experimentally validated, demonstrating robust performance across diverse environments and subjects [8]. Another significant novelty is Si-FI's hierarchical learning design, which combines subject identification and activity classification into a two-stage detection process. The cascaded architecture reduces overall complexity, transforming the exponential growth of multi-subject activity classification into a manageable problem with linear scalability.

The reliance of Si-FI on BFAs as its primary sensing primitive ensures seamless compatibility with commercial off-the-shelf Wi-Fi devices, eliminating the need for any firmware modification. While BFAs has been explored in recent Wi-Fi sensing works (as detailed in Section 3.2), Si-FI is the first work to leverage BFAs in simultaneous multi-subject sensing. This advancement enhances practicality and enables broader deployment across diverse environments. Achieving up to 98.94 % classification accuracy in real-world scenarios, Si-FI demonstrates its adaptability and sets a solid benchmark for simultaneous multi-subject sensing.

Si-FI incorporates a novel BFAs-based few-shot learning algorithm – FREL that allows rapid adaptation to new environments and subjects with minimal additional data (as little as 15 seconds per class). The FREL framework combines embedding learning with meta-learning to achieve superior domain adaptation performance with respect to existing approaches in dynamic and diverse environments. This integration not only ensures robust generalization but also delivers up to a 20 % improvement over state-of-the-art models [8,18] in unseen scenarios, demonstrating its adaptability and effectiveness. Moreover, Si-FI incorporates a lightweight and efficient preprocessing pipeline specifically designed for BFAs data alignment. This preprocessing enhances compatibility with IEEE 802.11-compliant devices.

3. Related work

Over the last ten years, a lot of efforts have been made to explore wireless sensing solutions, as summarized by Liu et al. in [28]. The reader may refer to the following surveys for a good compendium of the state of the art [5–7,29]. There have been several approaches to address the challenges of Wi-Fi sensing, which include received signal strength indicator (RSSI) based approaches [30,31], and passive Wi-Fi radar (PWR) [32,33]. More recently, researchers have focused on the more fine-grained CSI information that describes how the wireless channel modifies signals at different frequencies rather than providing a cumulative metric on the signal attenuation as the RSSI does [8,27]. Wi-Fi sensing leveraging beamforming feedback information (BFI) which can be captured from any Wi-Fi network is one of the state-of-the-art approaches [12,13]. However, CSI-based Wi-Fi sensing is by far the most popular approach.

Table 1 provides a comprehensive comparison between the proposed BFAs-based simultaneous multi-subject sensing (Si-FI) and other state of the art (SOTA) CSI- and BFI-based sensing approaches. The table includes information about the IEEE 802.11 standard utilized, the operational bandwidth, the considered sensing primitives, the need for any firmware modification, the addressed sensing tasks, the support for simultaneous multi-subject sensing together with the number of simultaneous subjects, the achieved sensing accuracy (based on the results in the reference papers), and the domain generalization capabilities. The comparison highlights the advantages of Si-FI with respect to other sensing approaches. Specifically, Si-FI performs BFAs-based simultaneous multi-subject sensing without requiring any firmware modifications, and allows achieving high accuracy across multiple tasks, offering superior generalization performance (achieving up to 99 % accuracy) across different scenarios. These factors make Si-FI the preferred choice for practical applications where simultaneous multi-subject sensing and robustness to domain shifts are critical. In the following, we summarize some of the main sensing approaches proposed in the literature considering both CSI- and BFI-based strategies.

3.1. CSI-based approaches

Deep learning (DL) has proven its effectiveness in various CSI-based Wi-Fi sensing applications, including tasks such as human activity classification [24,34–36], gesture recognition [23,25,37,38], health monitoring [39–41], human counting [42,43], and indoor localization [10]. Readers interested in further exploring DL-based CSI sensing

Table 1
Comparison between Si-FI and other state-of-the-art approaches for Wi-Fi sensing.

Reference	IEEE 802.11	Considered Bandwidth (MHz)	Sensing Primitive	Firmware Modification	Sensing Task	Sim. multi-sub. sensing	Sim. subjects	Sensing Accuracy (%)	Domain Generalization Accuracy
SiMWiSense [8]	AC	80 MHz	CSI	Yes	Activity classification	Yes	3	99.94	96.
MUSE-Fi[19]	AC	80 MHz	CSI, BFI	Yes (CSI extraction)	Activity and gesture recognition	Yes	8	97-99 (Gesture, Activity)	Not reported
Bi-directional BFM [20]	AX	80 MHz	BFI (EVD on autocorrelation matrix)	No	Human localization and AoD estimation	No	N/A	95-98	92-96
Respiratory Rate Estimation [21]	AX	80 MHz	BFI	No	Respiratory rate estimation	No	N/A	Error < 3.5 breaths/minute	Not reported
BFI-based sensing [22]	AX	80 MHz	BFI	No	Localization, tracking, and gesture recognition	No	N/A	92.5-97.14	Localization error: 0.3-0.72 m, Tracking error: 0.67-0.95 m
SignFi [23]	N	40 MHz	CSI	Yes	Sign gesture classification	No	N/A	94.81-98.91	86.66
WiAR [24]	N	40 MHz	CSI	Yes	Human activity recognition	No	N/A	80-95	80-90
OneFi [25]	N	40 MHz	CSI	Yes	Human gesture recognition via one-shot learning	No	N/A	84.2-98.8	75-91
BFMSense [26]	AC	80 MHz	BFI	No	Respiratory rate estimation and tracking	No	N/A	99 (Respiration)	Not reported
SHARP [27]	AC	80 MHz	CSI	Yes	Human activity recognition via micro-Doppler	No	N/A	>95	90-95
Si-FI (proposed)	AC	80 MHz	BFAs	No	Activity recognition	Yes	3	96-99	96-99

applications are referred to [29,44]. A few noteworthy works on human activity classification are briefly outlined below.

Shalaby et al. [34] proposed two DL models, namely a convolutional neural network (CNN)-Gated Recurrent Unit (GRU), and a CNN-GRU with attention, able to achieve an accuracy of 99.31 % and 99.16 % respectively in human activity classification. However, these models only consider a single subject performing six activities and can not generalize to new environments not considered during training. MCBCAR presented by Wang et al. [45] leverages generative adversarial network (GAN) and semi-supervised learning to address the challenge of dealing with non-uniformly distributed data due to environmental dynamics. Even though this work considers the dynamic change in the environment, the framework is not designed to adapt to new untrained environments and simultaneous multi-subject sensing. The AFSL-HAR framework by Wang et al. [46] uses few-shot learning to recognize new activities with a few samples of new data. Even though this work addressed the challenges of working with new scenarios, activities, or subjects through fast adaptation with few new samples, the framework classifies only one subject at a time in any environment. Ding et al. proposed RF-Net [47], a meta-learning framework to adapt to new environments with few labeled data samples. However, the RF-Net's CSI-based sensing performance ranges in (70-80)%.

SiMWiSense [8] is the first work to propose simultaneous multi-subject classification achieving up to 98.94 % of accuracy with baseline CNN whereas the few-shot based domain generalization algorithm FREL achieves an accuracy up to 96.53 % with new untrained environments/subjects. However, SiMWiSense is based on CSI which hampers the system scalability, overall system latency, and channel occupation as mentioned in Section 1. Hu et al. propose MUSE-Fi [19], a Wi-Fi-based multi-person sensing system leveraging near-field channel variations and sparse recovery algorithms to achieve physical separability and handle intermittent traffic in multi-user environments. The main idea is that devices that are close to the subjects to be sensed experience the higher variation in the CSI. Hence, MUSE-Fi assumes that users perform the activities close to the sensing devices. Instead, Si-FI achieves simultaneous multi-subject sensing without the reliance on near-field constraints. Moreover, as SiMWiSense, MUSE-Fi relies on CSI while Si-FI is fully based on compressed BFI data that allows reducing sensing latency and enhances the system scalability.

3.2. Recent BFI-based approaches

The utilization of BFI is gaining traction within the research community as a proxy for the CSI, offering a wealth of spatially diverse channel data from Wi-Fi-compliant devices without necessitating firmware modifications or any direct access to the device. For this, in [13], we presented Wi-BFI, one of the pioneering tools for obtaining the BFI from commercial devices in the form of BFAs, and reconstructing the beamforming feedback matrix from this data. A preliminary case study for activity recognition through BFI is also presented in [13]. As part of the BFI-based sensing research work in the literature, Jiang et al. illustrate how the formulation and computation of the beamforming feedback matrix influence sensing performance [48]. Kondo et al. assessed the effectiveness of uni-directional and bi-directional beamforming in Wi-Fi sensing using the beamforming feedback matrices reconstructed from BFAs [20]. The same authors also employed BFI for respiratory rate estimation [21]. Wu et al. proposed a BFI-based wireless sensing system for device localization, passive tracking, and sign language recognition [22]. Their proposed system achieves a localization median error of 0.72 m, passive tracking median error of 0.67-0.95 m, and sign language recognition accuracy of 92.5 % - 97.14 %. Yi et al. propose BFM-Sense [26], a framework that establishes the theoretical foundation for beamforming feedback-based Wi-Fi sensing, enabling fine-grained sensing.

However, all these studies focus on the *beamforming matrices reconstructed from BFAs*, which entails additional preprocessing, system latency, and computational overhead for the sensing system. Moreover, all these works are based on single-subject sensing. On the contrary, Si-FI focuses on simultaneous multi-subject sensing and is based on the compressed BFI which allows recording the BFAs of all the participating STAs with a single capture at the edge server. In this paper, the terms "compressed BFI" and "BFAs" are used interchangeably, as the quantization of BFI in IEEE 802.11 yields the BFAs transmitted by the STA to the AP as detailed in Section IV-A.

4. Overview of Si-FI

The architectural framework of Si-FI unfolds through three essential stages: firstly, the acquisition of BFI frames at the edge server; sec-

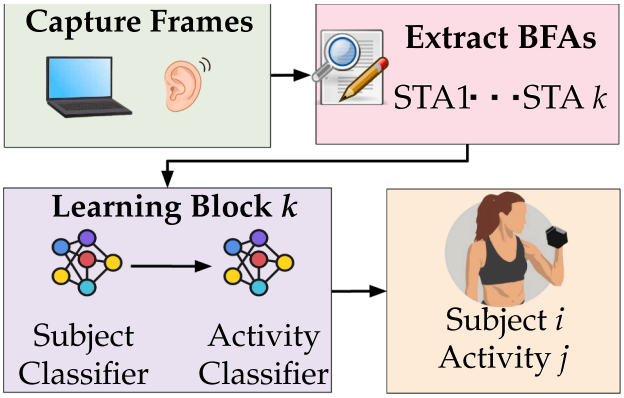


Fig. 2. Si-FI overview.

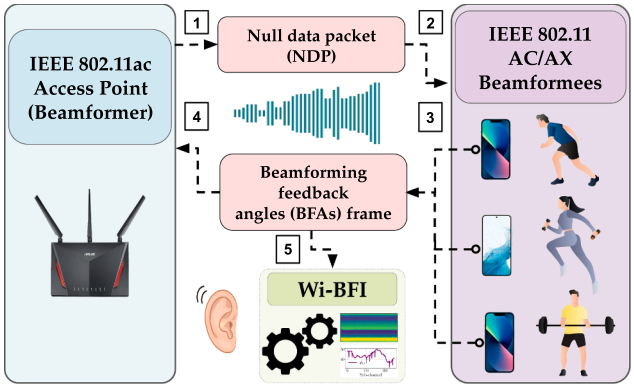


Fig. 3. BFI acquisition with Wi-BFI tool leveraging channel sounding in Wi-Fi systems..

ondly, the extraction of BFAs, and finally, the learning-based subject and activity classification. The pipeline is summarized in Fig. 2 and detailed in the following.

4.1. Acquisition of BFI frames

We leverage the Wi-BFI [13] tool for simultaneously capturing BFI frames from all the STAs close to one of the participating subjects. The BFI frames are continuously transmitted by the beamformees to the beamformer in a Wi-Fi network as part of the procedure implemented to set up a MIMO transmission.

Specifically, in MIMO systems, the beamformer *pre-codes* the data packets by linearly combining at the different transmitting antennas, the signals intended for simultaneous transmission to multiple beamformees. This process involves utilizing a precoding matrix $\mathbf{W}$ derived from the CFR matrix $\mathbf{H}$, which describes how the surrounding environment modifies the transmitted signals as they travel to the receiver. For this, each beamformee must estimate the CFR and provide feedback to the beamformer to ensure accurate precoding.

The process of estimating the CFR is termed channel sounding and is illustrated in Fig. 3. The beamformer initiates this procedure and involves periodically transmitting an null data packet (NDP) to gauge the MIMO channel between itself and the connected beamformees (depicted as **step 1** in Fig. 3). As their aim is to estimate the channel, the NDP frames are deliberately not subjected to beamforming. This approach offers a notable advantage: the resulting CFR estimation remains unaffected by either inter-stream or inter-user interference. The NDP comprises bit sequences, referred to as long training field (LTFs), whose decoded versions are known to the beamformees. These LTFs are transmitted over different beamformer antennas in consecutive time slots, enabling each beamformee to estimate the CFR of the links between its receiver antennas and the beamformer's transmitter antennas. As the other data fields, the LTFs are modulated through orthogonal frequency-division multiplexing (OFDM), i.e., they are transmitted through K partially overlapping and orthogonal sub-channels spaced apart by $1/T$.

Each beamformee in the system receives and decodes the NDP, as depicted in **step 2** in Fig. 3, to estimate the CFR indicated by the matrix $\mathbf{H}$. The estimation process utilizes various LTFs to estimate the channel characteristics for each combination of transmitter (TX) and receiver (RX) antennas across all OFDM sub-channels. In turn, $\mathbf{H}$ obtained at each beamformee is $K \times M \times N$ dimensional, where M and N denote the numbers of TX and RX antennas respectively. Subsequently, the CFR undergoes compression to mitigate channel overhead and is fed back to the beamformer. Specifically, by utilizing $\mathbf{H}_k$ to identify the $M \times N$ sub-matrix of $\mathbf{H}$ containing the CFR samples pertinent to sub-channel k, the *compressed beamforming feedback* is obtained as follows (see [49], Chapter 13). First, $\mathbf{H}_k$ is decomposed through singular value decomposition (SVD) as $\mathbf{H}_k^T = \mathbf{U}_k \mathbf{S}_k \mathbf{Z}_k^\dagger$, where $\mathbf{U}_k$ and $\mathbf{Z}_k$ are, respectively, $N \times N$ and $M \times M$ unitary matrices, while the singular values are collected in the $N \times M$ diagonal matrix $\mathbf{S}_k$. Using this decomposition, the complex-valued beamforming matrix $\mathbf{V}_k$ is defined by collecting the first $N_{\text{SS}} \leq N$ columns of $\mathbf{Z}_k$. Such a matrix is used by the beamformer to compute the pre-coding weights for the N_{SS} spatial streams directed to the beamformee. As part of the compression process, the $\mathbf{V}_k$ is then converted into polar coordinates, decomposing it as the product of multiple $\mathbf{D}_{k,i}$ and $\mathbf{G}_{k,\ell,i}$ matrices, which are defined as

$$\mathbf{D}_{k,i} = \begin{bmatrix} \mathbb{I}_{i-1} & 0 & & \cdots & 0 \\ 0 & e^{j\phi_{k,i,i}} & 0 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 & \\ & \vdots & 0 & e^{j\phi_{k,M-1,i}} & 0 \\ 0 & \cdots & & 0 & 1 \end{bmatrix}, \tag{1}$$

$$\mathbf{G}_{k,\ell,i} = \begin{bmatrix} \mathbb{I}_{i-1} & 0 & & \cdots & 0 \\ 0 & \cos\psi_{k,\ell,i} & 0 & \sin\psi_{k,\ell,i} & \vdots \\ \vdots & 0 & \mathbb{I}_{\ell-i-1} & 0 & \\ & -\sin\psi_{k,\ell,i} & 0 & \cos\psi_{k,\ell,i} & 0 \\ 0 & \cdots & & 0 & \mathbb{I}_{M-\ell} \end{bmatrix}, \tag{2}$$

where the values of the ϕ and ψ angles are obtained through the procedure detailed in Algorithm 1. The number of ϕ and ψ angles depends on the specific MIMO network configuration, i.e., the number of transmitter antennas and spatial streams. Using these matrices and $\tilde{\mathbf{D}}_k$ (see line 2 in Algorithm 1), $\mathbf{V}_k$ can be written as $\mathbf{V}_k = \tilde{\mathbf{V}}_k \tilde{\mathbf{D}}_k$, with

$$\tilde{\mathbf{V}}_k = \prod_{i=1}^{\min(N_{\text{SS}}, M-1)} \left(\mathbf{D}_{k,i} \prod_{l=i+1}^{M} \mathbf{G}_{k,l,i}^T \right) \mathbb{I}_{M \times N_{\text{SS}}}, \tag{3}$$

where the products represent matrix multiplications. In the $\tilde{\mathbf{V}}_k$ matrix, the last row – i.e., the feedback for the M-th transmitter antenna – consists of non-negative real numbers by construction. Using this transformation, the beamformee is only required to transmit the ϕ and ψ angles to the beamformer as they allow reconstructing $\tilde{\mathbf{V}}_k$ precisely. Moreover, it has been proved (see [49], Chapter 13) that the beamforming performance is equivalent at the beamformee when using $\mathbf{V}_k$ or $\tilde{\mathbf{V}}_k$ to construct the steering matrix $\mathbf{W}$. In turn, the feedback for $\tilde{\mathbf{D}}_k$ is not fed back to the beamformer.

To further reduce the channel occupation, the angles are quantized using b_ϕ bits for ϕ and $b_\psi = b_\phi - 2$ bits for ψ, as follows:

$$[\phi, \psi] = \left[\pi \left(\frac{1}{2^{b_\phi}} + \frac{q_\phi}{2^{b_\phi - 1}} \right), \left(\frac{1}{2^{b_\psi + 2}} + \frac{q_\psi}{2^{b_\psi + 1}} \right) \right]. \tag{4}$$

In IEEE 802.11ac/ax, $b_\phi = \{9, 7\}$ bits and $b_\phi = \{6, 4\}$ bits are used for MU-MIMO and SU-MIMO systems respectively. The quantized values – $q_\phi = \{0, \ldots, 2^{b_\phi} - 1\}$ and $q_\psi = \{0, \ldots, 2^{b_\psi} - 1\}$ – are packed into the compressed beamforming frame (**step 3** in Fig. 3) and such *BFAs frame* is transmitted to the beamformer (**step 4** in Fig. 3). The b_ϕ and b_ψ can be

Algorithm 1 $\mathbf{V}_k$ matrix decomposition.

Input: $\mathbf{V}_k$
$\tilde{\mathbf{D}}_k = \text{diag}(e^{j\angle[\mathbf{V}_k]_{M,1}}, \ldots, e^{j\angle[\mathbf{V}_k]_{M,N_{SS}}})$
$\mathbf{\Omega}_k = \mathbf{V}_k \tilde{\mathbf{D}}_k^{\dagger}$
for $i \leftarrow 1$ to $\min(N_{SS}, M-1)$ **do**
  $\phi_{k,\ell,i} = \angle[\mathbf{\Omega}_k]_{\ell,i}$ with $\ell = i, \ldots, M-1$
  Compute $\mathbf{D}_{k,i}$ through Eq. (1)
  $\mathbf{\Omega}_k \leftarrow \mathbf{D}_{k,i}^{\dagger} \mathbf{\Omega}_k$
  for $\ell \leftarrow i+1$ to M **do**
    $\psi_{k,\ell,i} = \arccos\left(\frac{[\mathbf{\Omega}_k]_{i,i}}{\sqrt{[\mathbf{\Omega}_k]_{i,i}^2 + [\mathbf{\Omega}_k]_{\ell,i}^2}}\right)$
    Compute $\mathbf{G}_{k,\ell,i}$ through Eq. (2)
    $\mathbf{\Omega}_k \leftarrow \mathbf{G}_{k,\ell,i} \mathbf{\Omega}_k$
  end for
end for

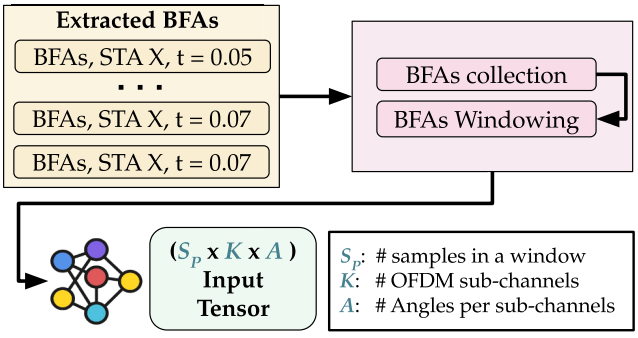


Fig. 4. BFAs data processing of each of the STAs in Si-FI.

read in the VHT/HE MIMO control field, together with other information like the number of columns (N_{SS}) and rows (M) in the beamforming matrix and the channel bandwidth. Each BFI contains A angles for each of the K OFDM sub-channels for a total of $A \times K$ angles each. We remark that, since MU-MIMO requires fine-grained channel sounding – around every 10 milliseconds, according to [50] – it is fundamental to process the BFI in a fast manner at the beamformer. For this reason, and since cryptography would lead to excessive delays, *the angles are currently transmitted over-the-air unencrypted*.

Wi-BFI[2] captures the frames and extract BFAs of each of the STAs associated with the participating subjects (**step 5** in Fig. 3). After the collection of the BFAs samples, we preprocess the captured data from each of the STA, as presented in Fig. 4, before feeding the data to the data-driven block. After S samples are collected by each of the STAs, we align the data by discarding the missing or corrupted BFAs measurements. Then, the captures are segmented into fixed-size non-overlapping windows along the time domain. This produces tensors of dimension of $S_p \times K \times A\times$ where S_p is the number of samples in p-th time window T_p. The tensors are then fed to the learning block.

4.2. Si-FI learning block

Scalability poses a significant challenge in multi-subject detection. With P individuals and Q activities, the potential combinations balloon is Q^P, leading to an exponential increase in class numbers when increasing the number of users. When P and Q are large, employing a single centralized model for classifying multi-subject activities becomes arduous. To address this issue, inspired by SiMWiSense [8], we propose a decentralized detection approach for each subject. Each device is assigned a learning model to detect the subject nearest to it. Consequently, only Q detection regions are necessary in hyperspace for each subject:

[2] https://github.com/kfoysalhaque/Wi-BFI

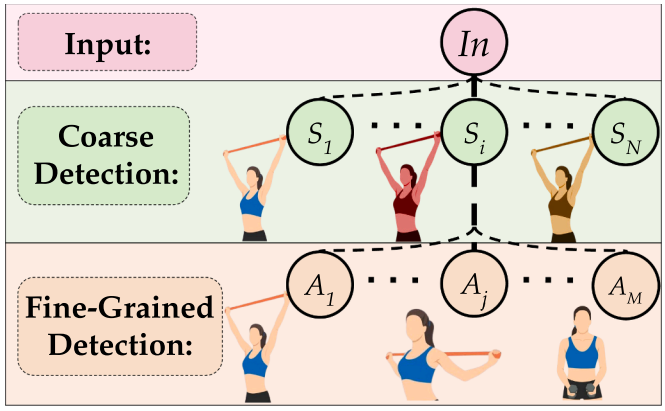


Fig. 5. Cascaded learning block.

With P subjects, the overall complexity diminishes to $P \times Q$. This system operates under two assumptions: firstly, there are at least as many STAs as subjects which is a plausible scenario given the ubiquitous daily use of smart devices like laptops and smartphones in both professional and personal contexts. Secondly, the subject closest to the device plays the most significant role in shaping the channel property between the device and the access point (AP) as proven in [8] through experimental analysis. This is reasonable because the path loss increases with the distance. Hence, the shorter the length of the path, i.e., the closer the users, the higher the impact on the CFR.

In Si-FI, the association between each sensing device and subject is determined and updated dynamically over time rather than through any fixed distance measurement. During both training and testing, every STA collects BFAs samples within each sensing interval (0.1 s) and performs coarse-grained subject identification to distinguish among the participating subjects and a "no-activity" state. Each STA then anchors itself to the subject whose movement induces the most dominant variation in its observed channel response, effectively identifying the nearest or most influential subject within its sensing region. As subjects move across the monitored area, the device that observes the strongest channel variation automatically re-associates itself with that subject in the subsequent sensing window. This adaptive, per-window reassessment enables seamless subject tracking and continuous activity recognition without requiring any explicit spatial calibration, handover, manual intervention, or additional signaling overhead. Such a design enhances the scalability and robustness of Si-FI in multi-subject environments by allowing autonomous and environment-aware STA–subject association.

To further reduce complexity, we introduce a cascaded model for two-stage detection. As depicted in Fig. 5, the first stage employs a DL model to coarsely distinguish different subjects S. Following the coarse detection, another DL model with finer granularity is utilized to identify the activities A. Notably, irrespective of the outcome in the initial stage, all subjects utilize the same fine-grained model in the second stage. Consequently, the overall complexity is reduced to $P + Q$. The hierarchical detection model faces a challenge: despite participating individuals performing the same activities, their movements exhibit unique patterns and gestures. Moreover, subjects may enter or exit the detection system, making a universal classifier for activity detection impractical. Furthermore, the performance of data-driven algorithms in wireless sensing often deteriorates due to fluctuating channel conditions over time. Therefore, there is a necessity for a model capable of rapid adaptation to new subjects and evolving channel features.

4.3. FREL learning algorithm

Few-Shot Learning (FSL), which emphasizes the rapid adaptation of deep neural network (DNN) to different tasks using only a small number of additional samples, is promising for addressing wireless tasks in dy-

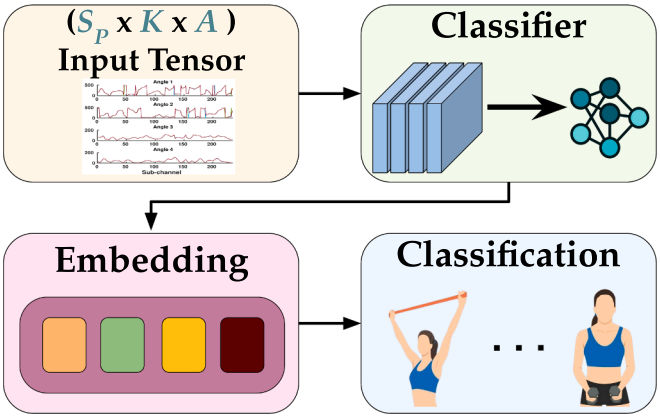


Fig. 6. Feature reusable embedding learning (FREL).

namic environments [8,51]. FSL methods can be broadly classified into embedding learning [52,53] and meta-learning [54,55]. In embedding learning, a DNN is utilized to learn a clustered feature space that can be reused for different tasks. During the inference phase, embedding learning requires a few labeled samples as anchors in the latent space to classify the new unlabeled data without the need for fine-tuning. On the other hand, meta-learning aims at learning a set of parameters that can be quickly optimized with a few additional data points in new environments, typically involving fine-tuning. In our previous study [8], we introduced Feature Reusable Embedding Learning (FREL) as a CSI sensing algorithm that combines embedding learning with meta-learning, demonstrating its capability to adapt to new scenarios with only a few CSI data samples. In this work, we integrate the Si-FI learning block with FREL, obtaining a FSL framework that can rapidly adapt to new BFAs data.

Specifically, an encoder-decoder structure is used in Si-FI FREL to jointly learn a reusable embedding and an adaptive classifier with fine-tuning, as illustrated in Fig. 6. Unlike the symmetric design adopted in auto-encoder designs, the encoder-decoder architecture in Si-FI FREL has an asymmetric design to integrate embedding learning and fast adaptation. The encoder, which has a larger computation capability compared to the decoder, is used to learn the generalized embedding that can be reused in different tasks. The decoder, on the other hand, is used as a classifier with a smaller architecture that can be fine-tuned rapidly. In parallel with meta-learning, Si-FI FREL involves both a meta-training and a fine-tuning phase. During the meta-learning phase, the encoder and decoder are jointly trained following the embedding learning principles. Once the embedding is obtained, only the classifier will be further fine-tuned using a small number of samples during testing. In contrast to embedding learning, fine-tuning offers greater flexibility and granularity, making it more suitable for dynamic systems. Different from meta-learning that fine-tune the whole DNN weights, Si-FI FREL only retrains the simple classifier, which can reduce the computation, enabling a faster adaption to new tasks. This design is inspired by [56], which indicates that the effectiveness of meta-learning primarily stems from feature reuse. It is a simplified version of MAML [54], fine-tuning the last few layers yet achieving compatible capability as the original algorithm.

Formally, let a functional mapping $E_\theta : X \to Z$ denote the Si-FI FREL encoder, where θ denotes the trainable parameter of the encoder while X and Z denotes the input and latent vector, respectively. The Si-FI FREL decoder can be modeled with another mapping $D_\phi : Z \to Y$, where ϕ and Y are the classifier parameters and the output label, respectively. Hence, The overall DNN architecture $F_\psi(X) = Y$ can be refined as $D_\phi(E_\theta(X)) = Y$, where $\psi = \{\theta, \phi\}$ is the total trainable parameters of the whole system. In FSL, models are trained with mini-batches of data sampled from different tasks, each consisting of N classes and K samples per class, referred to as "N-way-K-shot" learning. To incorporate FSL to Si-FI FREL, we model each batch of BFAs data as a new task $\tau_j = \{(x_i^j, y_i^j)\}|_{i=1}^m$, where $m = N \times K$ denotes the total number of samples in one batch. The objective of Si-FI FREL is to find the optimal parameters ψ that minimize the expectation of the loss function $\mathcal{L}$ with respect to a group of meta-learning tasks $\mathcal{T} = \{\tau_j\}|_{j=1}^t$, i.e.,

$$\min_{\{\theta,\phi\}} \quad \frac{1}{t}\sum_{j=1}^{t}\left[\frac{1}{m}\sum_{i=1}^{m}\mathcal{L}\left(C_\phi\left(E_\theta\left(x_i^j\right)\right) = y_i^j\right)\right] \tag{5}$$

In Eq. (5), the optimization problem involves minimizing the average loss over m samples for each of the t tasks. The expression optimizes the loss function jointly across both the samples and the tasks, without explicitly using inner or outer loops. Inspired by [56], Si-FI FREL merges all tasks $\mathcal{T}$ into a single dataset $\mathbb{D}^{train}$, thereby facilitating embedding learning. This is expressed as

$$\begin{aligned}\mathbb{D}^{train} &= \tau_1 \cup \cdots \cup \tau_j \cup \cdots \cup \tau_t \\ &= \{(x_i^1, y_i^1)\}|_{i=1}^m \cup \cdots \cup \{(x_i^n, y_i^n)\}|_{i=1}^m\end{aligned} \tag{6}$$

As a result, the two loops of optimization in Eq. (5) can be reduced to a general DL problem. This enables Si-FI FREL to solve the optimization problem using a stochastic gradient descent approach, iteratively updating the θ and ϕ parameters as

$$\{\theta, \phi\} = \{\theta, \phi\} - \alpha\frac{1}{mt}\sum_{i=1}^{mt}\nabla_{\{\theta,\phi\}}\mathcal{L}(D_\phi(E_\theta(x_i)), y_i) \tag{7}$$

where mt is the total number of data points in $\mathbb{D}^{train}$ and α is the learning rate.

While Eq. (7) allows Si-FI FREL to learn the optimal encoder E_{θ^*} to map input to a clustered embedding space on the task set $\mathcal{T}$, the parameters ϕ^* may not be optimal for new tasks. To address this issue, another small portion of new data $\mathbb{D}^{tune}$ is used to fine-tune the classifier D_ϕ after meta-learning. Unlike training on a combined set in Eq. (6), each iteration we randomly sample K shots from each of N ways in $\mathbb{D}^{tune}$ to build a new task τ and update classifier by gradient descent,

$$\phi = \phi - \beta\frac{1}{m}\sum_{i=1}^{m}\nabla_\phi\mathcal{L}(D_\phi(E_{\theta^*}(x_i)), y_i) \tag{8}$$

where β denotes the learning rate in the fine-tuning phase. The performance of the algorithm is evaluated on the rest of unseen dataset $\mathbb{D}^{test}$. The complete Si-FI FREL procedure is summarized in Algorithm 2.

In practical deployment, Si-FI adapts to a new environment or a new subject using only a small number of samples collected locally. During this process, the encoder parameters remain fixed, and only the lightweight classifier is fine-tuned following the 5-shot adaptation protocol described in Algorithm 2. The fine-tuning phase requires approximately 15 s of BFA data per activity class and completes within a few seconds on an edge GPU, enabling rapid model calibration without full retraining. This design allows Si-FI to maintain consistent performance across unseen environments and subject variations with minimal data and computational effort. Next, we discuss the specific setup that is used in the Si-FI learning block.

- **Embedding network:** The encoder of Si-FI FREL is referred to as "embedding network". As depicted in Fig. 7, the whole architecture consists of four convolutional layers (referred as to "ConvLayer" in Fig. 7), each convolutional layer consist of a 3×3 convolutional kernel with 64 channels (referred as to "Conv2D" in Fig. 7), a batch normalization ("BatchNorm"), and a Rectified Linear Unit (ReLU) activation function. Following the first three convolutional layers, 2×2 max pooling layers ("MaxPool") are employed to downsample the output of the preceding layer. Subsequent to the fourth convolutional layer, a global average pooling strategy ("GlobalAvg") is adopted to extract features into the latent space, yielding a 64-dimensional feature space.
- **Classifier:** A fully-connected layer is used on top of the embedding network as a linear decoder. Note that Si-FI does not involve complex

Algorithm 2 Feature reusable embedding learning.

Phase 1: FREL meta-learning
Require: learning rate α, dataset $\mathbb{D}^{train}$
Initialize: θ for embedding, ϕ for classifier
for $iteration = 1, 2, \ldots$ **do**
  Update θ and ϕ with $\mathbb{D}^{train}$ by Equation (7)
end for
Return: θ^* for embedding

Phase 2: FREL fine-tuning
Require: learning rate β, dataset $\mathbb{D}^{tune}$
Initialize: ϕ for classifier
for $epoch = 1, 2, \ldots$ **do**
  for $episode = 1, 2, \ldots$ **do**
    Sample a task $\tau = \{(x_i, y_i)|_{i=1}^{m}\}$ from $\mathbb{D}^{tune}$
    Update ϕ with τ by Equation (8)
  end for
end for
Return: ϕ^* for classifier

decoder designs since the aim is to investigate the efficacy of FREL algorithm on BFAs data. As a baseline comparison, an embedding learning algorithm [18] is deployed and trained with the same data set $\mathbb{D}^{train}$. In [18], an untrainable K-Nearest Neighbor (K-NN) classifier is applied during inference time without fine-tuning. K samples from each class in $\mathbb{D}^{tune}$ are mapped into embedding through the trained embedding network E_{θ^*} as supports, and the test data is classified by a plurality vote of the K nearest supports. We refer to this approach as embedding learning.

• **Dataset:** One major difference between the Si-FI FREL and general FSL lies in the dataset. General FSL datasets such as Omniglot [57] and Mini-ImageNet [52] contain a closed set of tasks. While additional samples are used for adaptation, the tasks performed as part of the pre-training and adaptation phases remain the same. Tasks in wireless sensing are considered distinct in different environments due to variations in the pattern of sensing primitives, which are influenced by changes in physical dynamics. Consequently, new tasks frequently emerge when transitioning between environments or when significant changes occur in the physical conditions of the same environment. This makes our problem an open set of challenges. Given this, to evaluate the generalization capabilities of our algorithm to the new tasks, the dataset in Si-FI is divided into a training set $\mathbb{D}^{train}$, a fine-tuning set $\mathbb{D}^{tune}$ and a test set $\mathbb{D}^{test}$. $\mathbb{D}^{train}$ contains a series of BFAs data collected in same environments. $\mathbb{D}^{tune}$ is a mini-dataset containing data collected in different environments for 15 seconds, while $\mathbb{D}^{test}$ entails BFAs samples collected in new environments never exposed to the model during pre-training and fine-tuning. As environments are time-varying, $\mathbb{D}^{tune}$ and $\mathbb{D}^{test}$ represent not only different samples but also new tasks, making the problem more challenging.

• **Learning strategy:** Si-FI applies 5-shot learning to FREL, sampling 5 data points for each class in every batch during pre-training and fine-tuning. Adam is used as the optimizer in both phases while both learning rates α and β are set to 0.01. The cross-entropy loss is considered for error backpropagation.

5. Experimental setup

We evaluate the multi-subject sensing capability of Si-FI as well as the generalizing feature of Si-FI FREL through an extensive data collection campaign in 3 different environments: classroom, office, and kitchen with 3 human subjects performing 20 different activities in random order. We used off-the-shelf Netgear Nighthawk X4S AC2600 routers to set up the network, whereas IEEE 802.11ac compliant Asus RT-AC86U routers empowered with the Nexmon tool were used as for CSI extraction [16]. We evaluated the multi-subject sensing capability of Si-FI and the generalization power of Si-FI FREL through a comprehensive data collection campaign conducted across three distinct environments: classroom, office, and kitchen. Three human subjects performing 20 different activities in random order were involved. We collected the BFAs and the corresponding CSI for comparative analysis of the BFAs (Si-FI) and CSI (SiMWiSense) based approaches.

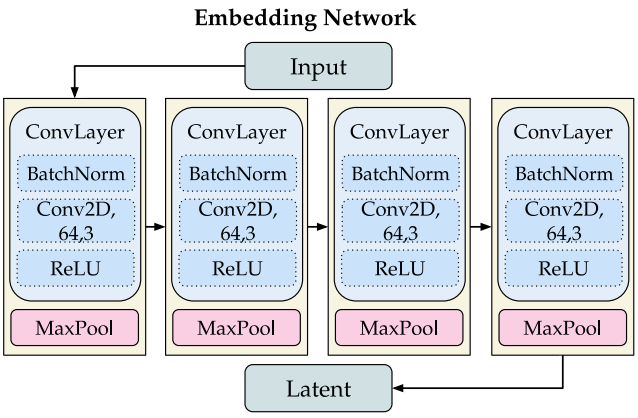


Fig. 7. CNN embedding network.

5.1. Si-FI network setup and equipment.

Our experimental network is based on an 802.11ac MU-MIMO framework, operating on channel 153 with a center frequency of 5.77 GHz and an 80 MHz bandwidth (256 sub-channels in total). Note that this configuration enables the sounding of 234 sub-channels (data sub-channels) as the channel is not estimated on the 14 control sub-channels and 8 pilot sub-channels during the sounding procedure. Our network infrastructure includes one access point AP functioning as a beamformer and three STAs acting as beamformees, depicted in Fig. 8 with the router picture and the red boxes respectively. Each device in the setup, both the AP and the STAs, is equipped with a Netgear Nighthawk X4S AC2600 router. The AP operates with three antennas ($M = 3$), while a single antenna is enabled on each STA ($N = 1$).

The three STAs receive one spatial stream each ($N_{ss} = 1$) to establish a 3×3 MU-MIMO system. According to the IEEE 802.11ac standard, four beamforming feedback angles are required (two for each of the angle groups ϕ and ψ) to characterize the 3×1 channels between the AP and the STAs. The feedback mechanism for these angles utilizes a 9-bit quantization for ϕ ($b_\phi = 9$) and a 7-bit quantization for ψ ($b_\psi = 7$).

5.2. CSI network setup and equipment.

Alongside our BFI frame capture, we concurrently collected CSI data for comparative analysis. This data was captured using three CSI monitors, co-located with the Si-FI STAs, as illustrated in green in Fig. 8. IEEE 802.11ac-compliant Asus RT-AC86U routers equipped with the Nexmon CSI extraction tool are used as CSI monitors [16]. To perform a fair comparison with the Si-FI network setup, each CSI monitor is configured to utilize one antenna ($N = 1$) and one spatial stream ($N_{ss} = 1$) to sense the channel.

The co-located STAs and CSI monitors are positioned within each environment maintaining a separation of 1.5 m to 2.0 m from each other, as depicted in Fig. 8. In each setting, three individuals perform a series of 20 activities, including pushing forward, rotating, hands up and down, waving, brushing, clapping, sitting, eating, drinking, kicking, bending forward, washing hands, making a call, browsing a phone, checking wrist, reading, waving while sitting, writing, side bending, and standing. These activities are executed at a distance of 1.5 m to 2.0 m from

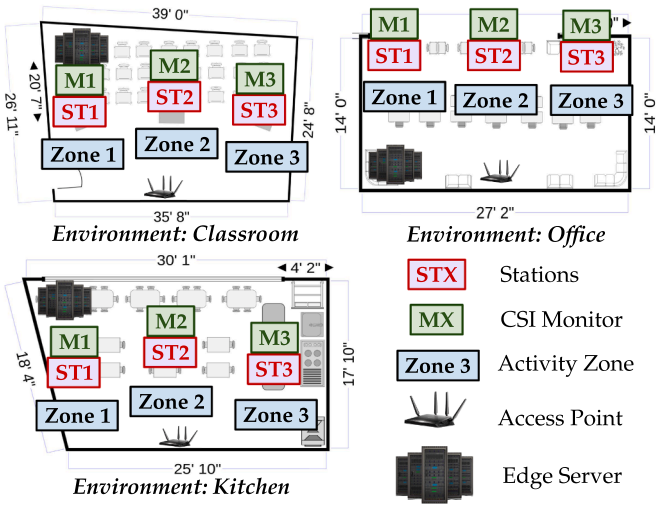


Fig. 8. Experimental setup of `Si-FI`.

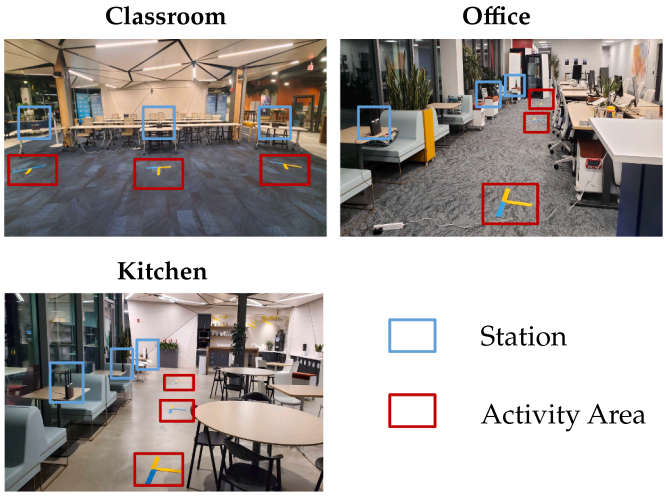


Fig. 9. Data collection locations.

the STA closest to the user, ensuring consistent and clear data capture across the experiments.

UDP data streams are transmitted from the AP to the STAs in the downlink direction, facilitating the initiation of BFAs and CSI collection. The BFAs frames are captured using the `Wi-BFI` tool, which is run on a commercially available laptop equipped with an Intel 9560NGW wireless-AC network interface card (NIC), configured in monitor mode. It is important to note that the laptop used for capturing these frames does not require a direct connection to either the AP or the STAs. The primary requirement is that the frame capture occurs on the same wireless channel on which the Wi-Fi network operates. The `Wi-BFI` tool extracts the ϕ and ψ angles for each of the STAs, which are then fed into the `Si-FI` learning framework (refer to Section 4.2). Concurrently, each co-located CSI monitor, equipped with the Nexmon tool, is triggered by the UDP transmissions and captures the corresponding CSI data.

Three distinct models are trained for each of the STAs with the BFAs data from that STA while the subject closest to it performs activities while the other subjects execute random movements. For instance, the model for STA 1 is trained with Subject 1 executing each of the 20 activities for 5 minutes, while Subject 2 and Subject 3 perform random movements (distinct from the aforementioned 20 activities) in random order. This approach ensures that each model captures the specific features associated with the monitored subject while suppressing the variability introduced by the presence of other subjects performing unrelated movements. The experimental setup and the activity areas in three experimental sites are depicted in Fig. 9.

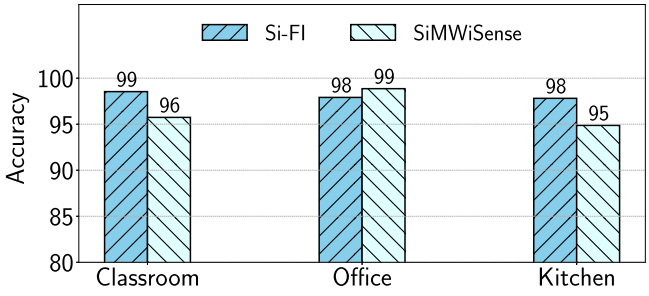
6. Performance analysis of `Si-FI`

The collected data is split into time windows of 0.1 s which results in 10 samples per window. Initially, we evaluate the performance of our two-stage detection system, encompassing (i) subject identification (coarse-grained detection) and (ii) activity classification (fine-grained detection), utilizing the baseline CNN architecture depicted in Fig. 7. Subsequently, we illustrate the generalization capability of `Si-FI FREL` across both stages of the detection system and compare its performance with both BFAs-based and state-of-the-art CSI-based approaches.

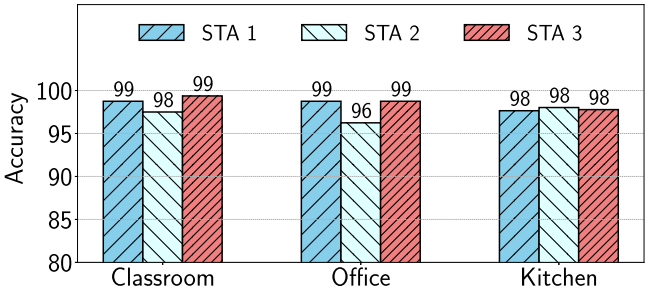
6.1. *Si-FI performance with baseline CNN*

6.1.1. *Subject identification performance*

In the first step of the two-stage `Si-FI` detection, each STA classify Sub1, Sub2, Sub3 or "no activity". The subject identification performances of both `Si-FI` and `SiMWiSense` for each of the STAs in three environments are presented in Fig. 10(a).



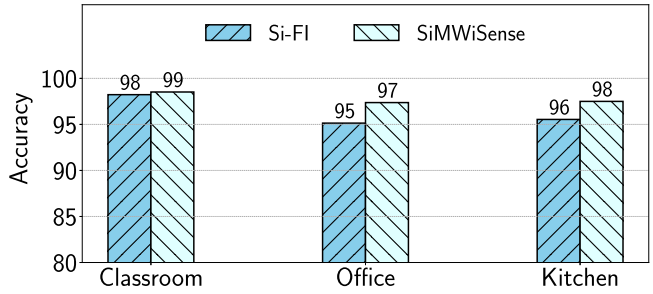
(a) BFAs-based approach (`Si-FI`) vs CSI-based approach (`SiMWiSense`). The results are averaged over different subjects and STAs.



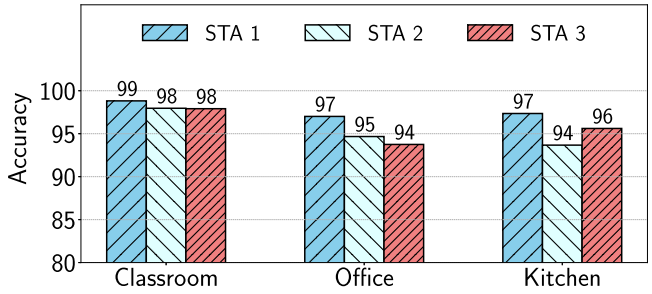
(b) `Si-FI` subject identification (coarse-grained detection) performance across different STAs. The results are averaged over different subjects.

Fig. 10. Subject identification (coarse-grained detection) performance in different environments using the baseline CNN.

The results show that with the baseline CNN, the average subject identification accuracy of `Si-FI` in classroom, office, and kitchen across all three STAs are 98.76 %, 97.72 % and 98.43 % respectively, showing no significant discrepancy among the different environments. The performance of `SiMWiSense` across the three environments is 96.42 %, 99.32 %, and 94.88 % respectively. Hence, overall, for subject identification, `Si-FI` achieves a classification accuracy comparable to the one of the corresponding CSI-based approach `SiMWiSense`. Fig. 10(b) presents subject identification performance with the data from different



(a) BFAs-based approach (`Si-FI`) vs CSI-based approach (`SiMWiSense`). The results are averaged over different subjects and STAs.



(b) `Si-FI` activity classification performance of all the STAs across all environment. The results are averaged over different subjects.

Fig. 11. Activity classification (fine-grained detection) performance with baseline CNN.

STAs for each of the considered environments. The results show that, in the classroom and kitchen, different STAs do not show significant performance discrepancy whereas, in the office, the performance of STA 2 is 3% lower than that of the other two STAs. It might be caused by the floor layout and poor channel estimation because of noise and interference.

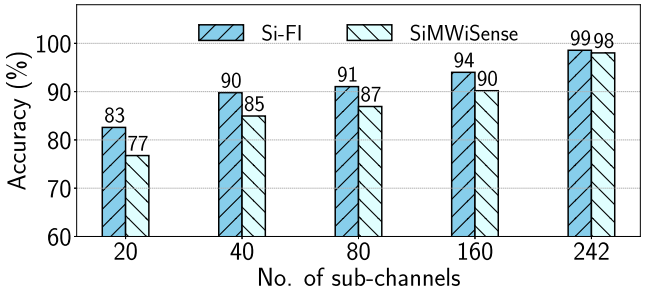
6.1.2. *Activity classification performance*

The simultaneous multi-subject activity classification performance of `Si-FI` with baseline CNN is presented in Fig. 11(a). The average accuracy in the office, classroom and kitchen are 98.22%, 95.44%, and 95.88% respectively, which follows the similar trend of subject detection performance depicting the stability and robustness of `Si-FI`. `SiMWiSense` achieved an average accuracy of 98% which is only 1.51% more than that of the BFAs-based approach `Si-FI`. Thus, also for the simultaneous activity classification task, `Si-FI` performance is similar to the CSI-based approach `SiMWiSense`. Fig. 11(b) shows that for activity classification (fine-grained detection) as well, the performance variation of the models across different STAs and environments remains minimal, averaging less than 2%. This consistency highlights the stability and robustness of the system.

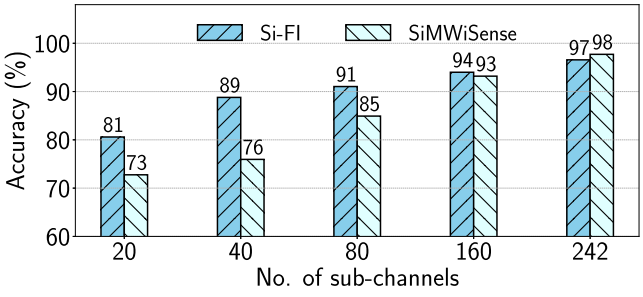
6.1.3. *`Si-FI` performance as a function of the bandwidth.*

It is known that Wi-Fi sensing performs worse with lower bandwidths (lower number of sub-channels) as the sensing resolution decreases [4,58]. To compensate for the lower resolution, one can adopt extensive feature extraction techniques which would increase the computation burden by intensifying the pre-processing steps and learning process dramatically. This stimulates us to study a trade-off between the number of sub-channels used for sensing and the `Si-FI` performance.

Fig. 12(a) and (b) present the performance of subject identification and simultaneous activity classification respectively as a function

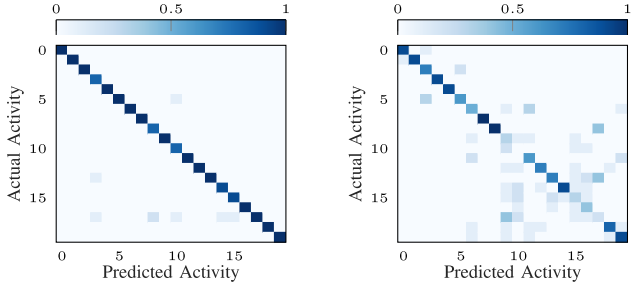


(a) Subject identification.



(b) Simultaneous activity classification.

Fig. 12. Average performance of `Si-FI` vs SiMWiSense with baseline CNN as a function of number of subcarriers.



(a) 242 sub-carriers, accuracy: 98.23%

(b) 20 sub-carriers, accuracy: 70.76%

Fig. 13. Confusion matrices of baseline CNN with 242 and 20 subcarriers respectively (in classroom, with monitor M2).

of the number of sub-channels. The first consecutive 20, 40, 80, 160, and 234 sub-channels are considered to emulate sensing systems with subsequently lower bandwidths. The results show that the average `Si-FI` performance of subject identification and activity classification decreases to 91.38% and 90.76% respectively when we switch to 80 sub-channels and 83.18% and 81.34% in the case of 20 sub-channels only. `SiMWiSense` also shows a similar trend when reducing the number of sub-channels considered for sensing. Fig. 13 presents the confusion matrices of the baseline CNN for simultaneous activity classification with `Si-FI` when trained using 242 and 20 sub-channels, respectively. The results are shown for the classroom environment, which is used as a representative example since similar behavior was consistently observed in the office and kitchen setups. When trained with all 242 sub-channels, the model achieves high accuracy across almost all activity classes, indicating that the CNN is able to capture fine-grained variations in the beamforming feedback (BFI/CSI) caused by user motion. When the number of sub-channels used in the classification is reduced to 20, certain classes such as *wash hands* (index 17), *rotate* (index 10), and *brush* (index 15) experience noticeable performance

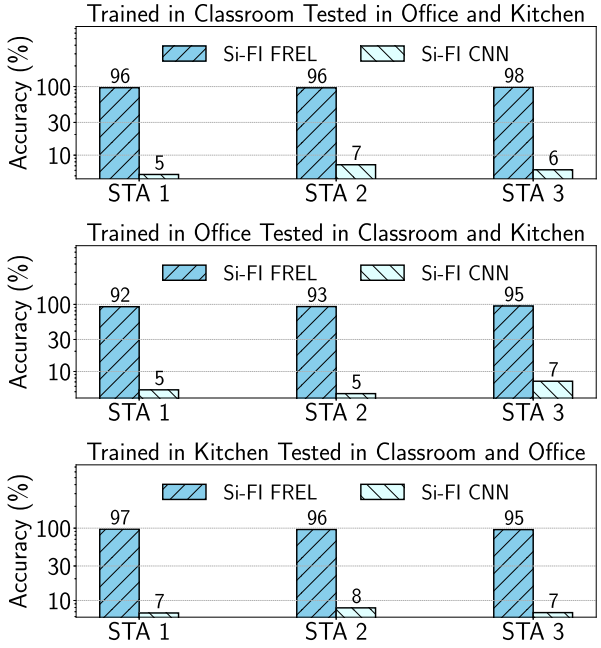


Fig. 14. Performance of FREL as the subject identifier in untrained environments.

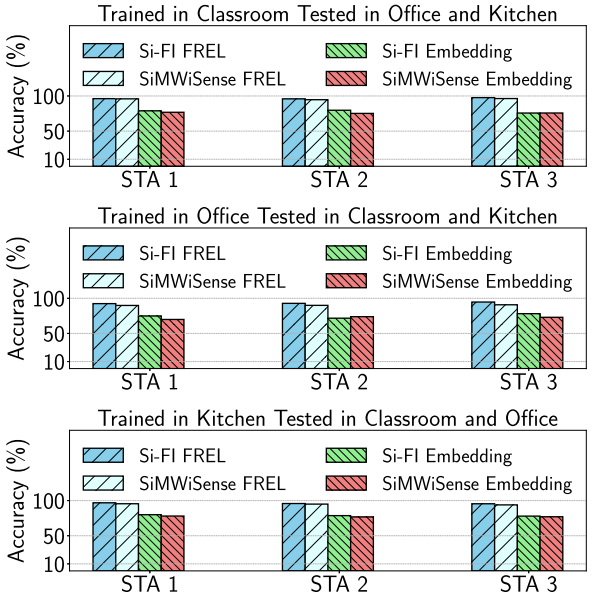


Fig. 15. Performance comparison of FREL and embedding learning as the subject identifier in untrained environments.

degradation. These activities involve localized or repetitive movements that induce only subtle phase and amplitude variations in the channel response, which become less distinguishable to the CNN as the frequency resolution decreases. The reduction from $50 \times 242 \times 2 = 24,200$ to $50 \times 20 \times 2 = 2,000$ input elements significantly limits the available frequency-domain information, thereby lowering separability between fine-grained motion patterns. Nevertheless, even under this reduced-bandwidth configuration, Si-FI maintains an average accuracy of approximately 73 %, which demonstrates its robustness to coarse spectral sampling.

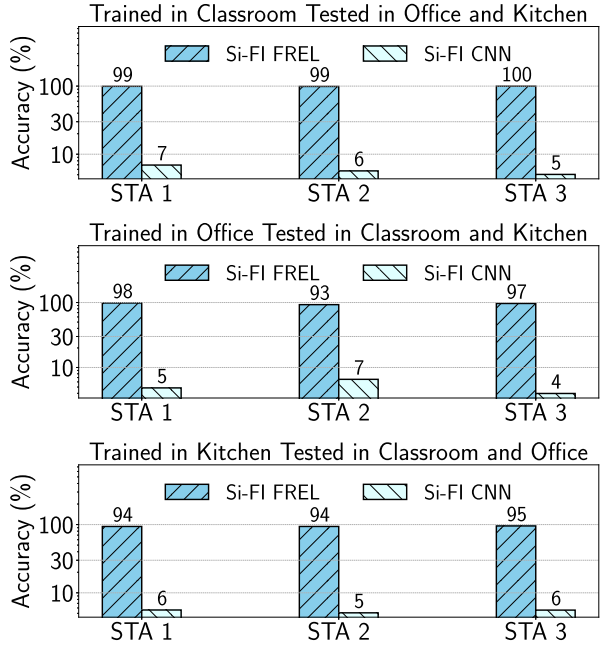


Fig. 16. Performance of Si-FI FREL in simultaneous activity sensing on new untrained environments.

6.2. *Generalization performance of Si-FI*

Even though the performance of the traditional CNN is quite good, they fail to generalize the environments or subjects that are not part of the training dataset. In generalizing to new environments and subjects, Si-FI FREL excels in comparison to the traditional CNN. For comparative analysis, we also implemented Si-FI Embedding, an embedding learning algorithm [18] for Si-FI. We also compare the Si-FI generalization performance with the CSI-based simultaneous sensing approaches SiMWiSense FREL and SiMWiSense Embedding.

6.2.1. *Si-FI FREL performance in generalizing subject identification across environments*

Fig. 14 presents the comparative analysis of subject identification performance of Si-FI FREL with baseline CNN whereas Fig. 15 compares the performance of Si-FI FREL with the earlier mentioned state-of-the-art algorithms, i.e., Si-FI Embedding (BFAs), SiMWiSense FREL (CSI), and SiMWiSense Embedding (CSI). The results show that the overall performance of Si-FI FREL in new untrained environments improves by 89.44 %, 84.18 %, and 87.65 % when the embedding network is trained in the classroom, office, and kitchen environments respectively in comparison to the baseline CNN. Thus, the basic CNN cannot provide generalization capabilities for the subject identification task. On the contrary, the proposed Si-FI FREL performs well even in new untrained environments, as presented in Fig. 15: the average highest performance of 95.30 % is achieved by Si-FI FREL, whereas, Si-FI Embedding (BFAs), SiMWiSense FREL (CSI), and SiMWiSense Embedding (CSI) achieve an average accuracy of 77.26 %, 93.52 %, and 75.13 % respectively across different environments and STAs. Specifically, Si-FI FREL improves the performance by 1.86 % and 21.16 % respectively in comparison to CSI-based approaches (SiMWiSense FREL and SiMWiSense Embedding) whereas it surpasses the BFAs-based approach (Si-FI Embedding) by 18.92 % on an average across different STAs and environments.

6.2.2. *Si-FI FREL performance in generalizing activity classification across environments*

Si-FI FREL performs well – with an average accuracy of 96.52 % – in generalizing the simultaneous activity classification on new untrained

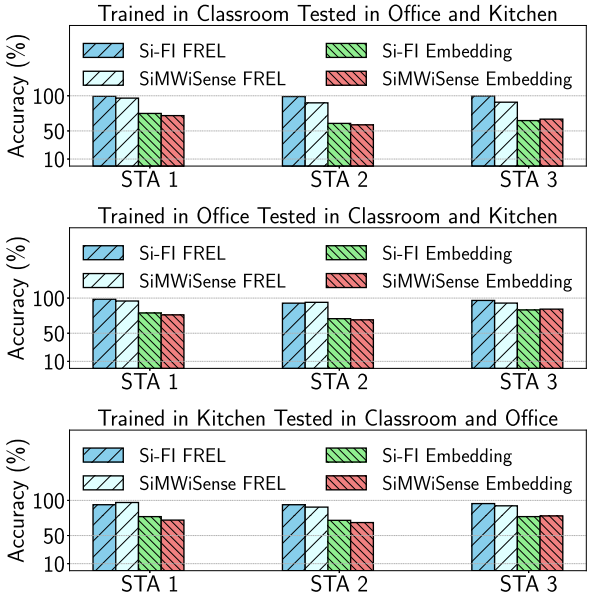


Fig. 17. Performance comparison of `Si-FI FREL` with SiMWiSense-`FREL` and state-of-the-art embedding learning based approaches in adapting new untrained environments.

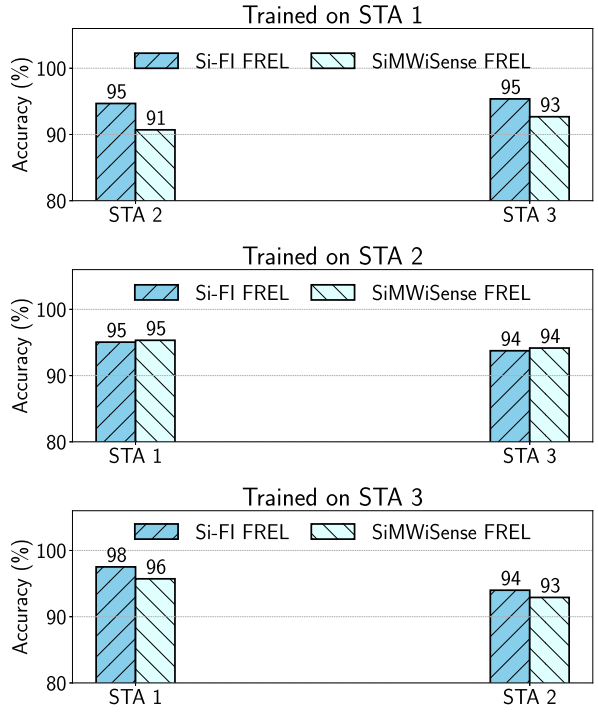


Fig. 18. Performance of `Si-FI FREL` with new untrained monitors.

environments, as shown in Figs. 16 and 17. Specifically, `Si-FI FREL` can achieve up to an accuracy of 99.67 % when trained in the classroom and tested in the office and kitchen whereas the traditional CNN reaches only 6.19 % across different STAs and environments. As depicted in Fig. 17, when compared to `Si-FI Embedding` (BFAs), `SiMWiSense FREL` (CSI), and `SiMWiSense Embedding` (CSI), the proposed `Si-FI FREL` demonstrates an average accuracy boost of 24.21 %, 3.32 % and 25.75 % respectively. Moreover, the accuracy of `Si-FI FREL` fluctuates only by

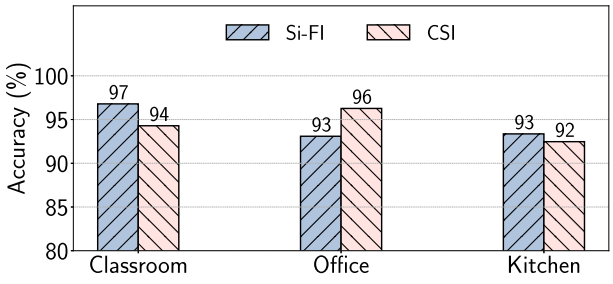


Fig. 19. Joint subject–activity recognition performance of the baseline CNN model using BFI (`Si-FI`) and CSI-based (`SiMWiSense`) approaches.

2.05 % and 4.08 % on average across different STAs and environments. In turn, `Si-FI FREL` provides stable and reliable performances across the STAs and environments in performing the activity classification task.

6.2.3. *`Si-FI FREL` performance in generalizing activity classification across stations*

Here we investigate whether the `Si-FI FREL` model trained in one STA can adapt and generalize to other STAs. This would allow further reducing the training time thus drastically reducing the system deployment complexity. We compare simultaneous activity classification performance of the two highest performing schemes, i.e., `Si-FI FREL` (proposed) and `SiMWiSense FREL` (CSI), while trained with the data from one STA and tested with data collected from other STAs as shown in Fig. 18. `Si-FI FREL` achieves an average accuracy of 95.02 %, 94.4 %, and 95.76 % while trained with the data from STA 1, STA 2, and STA 3 respectively and tested with the others. On the other hand, the state-of-the-art CSI-based approach – `SiMWiSense FREL`– achieves an average accuracy of 93.58 % which is comparable to `Si-FI FREL` (the proposed BFAs-based approach). Thus `Si-FI FREL` enables the sensing system to be trained with data from only one STA and deployed with an arbitrary number of STAs using only 15 s of new data samples from the untrained STAs.

6.3. *Joint subject–activity evaluation*

To obtain a unified perspective on end-to-end sensing performance, we further evaluate `Si-FI` using a joint metric that simultaneously considers both subject identification and activity classification. A prediction for a sample is considered correct only when both the subject and the associated activity are correctly identified.

In the baseline CNN evaluation (Fig. 19), `Si-FI` achieves joint accuracies of 96.79 %, 93.09 %, and 93.36 % in the classroom, office, and kitchen environments, respectively. On average, `Si-FI` improves the joint recognition performance by approximately 1.8 % over the CSI-based baseline in environments with stronger multipath variations (classroom and kitchen). A similar trend is observed in the fine-tuned FREL models (Fig. 20). The proposed `Si-FI`-FREL architecture achieves the highest joint accuracies, reaching 97.34 % in the kitchen and maintaining 95.44 % and 94.76 % in the classroom and office, respectively, representing an average gain of 6–9 % compared to `SiMWiSense`-FREL. The gain becomes even more pronounced when compared to embedding-only variants, whose joint accuracy drops below 60 %. By minimizing cumulative misclassification across the two inference stages, `Si-FI` sustains robust subject-aware sensing performance across heterogeneous environments, confirming its scalability and practicality for real-world deployment.

6.4. *Channel occupation analysis of `Si-FI`*

Given the limited availability of communication resources and the time-sensitive nature of sensing applications, the efficiency of any

Table 2
Comparison of latency components of different approaches when varying the number of sub-channels.

no. of sub-channels	Approaches	Latency components			
		Tx time (ms)	Extraction time (ms)	Inference time (ms)	End-to-end latency (ms)
234	**Si-FI**	**0**	**10.7**	**20.2**	**30.9**
	SiMWiSense	11	54	18.2	83.2
160	**Si-FI**	**0**	**7.5**	**19.1**	**26.6**
	SiMWiSense	6	38	17.3	61.3
80	**Si-FI**	**0**	**6.5**	**16**	**22.5**
	SiMWiSense	3.5	28	15.7	47.2
40	**Si-FI**	**0**	**5.5**	**15.3**	**20.8**
	SiMWiSense	3	22	14.3	39.3
20	**Si-FI**	**0**	**5**	**15.1**	**20.1**
	SiMWiSense	3	19	14	36

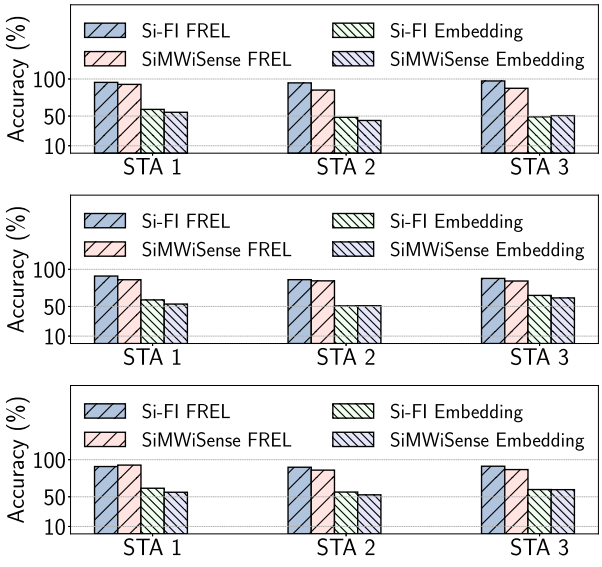


Fig. 20. Joint subject–activity performance comparison of `Si-FI`-FREL with `SiMWiSense`-FREL and state-of-the-art embedding learning based approaches in adapting new unseen environments.

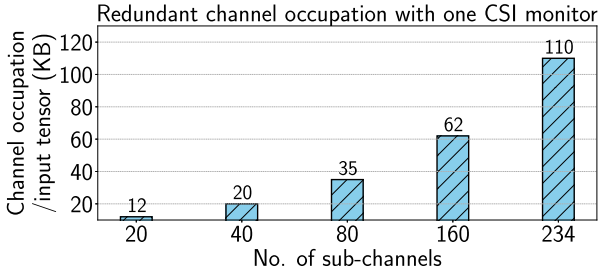


Fig. 21. Redundant channel occupation per input tensor for a single `SiMWiSense` CSI monitor with different number of considered sub-channels.

sensing system hinges on its channel occupancy and end-to-end latency performance. For this, in our system, the BFAs are directly acquired by `Wi-BFI` at the server, eliminating the need for edge offloading and thereby minimizing channel occupancy. In essence, `Si-FI` operates without occupying any redundant channel space, regardless of the number of sensing STAs, transmission bandwidth, and considered number of sub-channels. On the other hand, CSI-based simultaneous multi-subject sensing methods introduce additional channel occupation overhead as, unlike `Si-FI`, data is captured at each CSI monitor and needs to be offloaded to the edge server. Thus, the overall channel occupation of the CSI-based approaches also depends on the total number of simultaneous sensing monitors. Notice that, the amount of data required to capture CSI for a given bandwidth and time window may vary slightly depending on the capturing tool utilized. Fig. 21 illustrates the channel occupation overhead of a `SiMWiSense` (CSI) input tensor (refer to Fig. 4) for a single CSI monitor operating on a channel with 80 MHz bandwidth. The results indicate that even a single input tensor captured within a 0.1-second time window occupies 110 KB when considering 234 sub-channels, which reduces to 35 KB for 80 sub-channels. While `Si-FI` entails zero-redundant channel occupancy, the overhead increases exponentially with CSI-based approaches like `SiMWiSense`. In a system with three simultaneous sensing monitors (as detailed in Section 5), the redundant channel occupancy per second for a channel bandwidth of 80 MHz and 234 sub-channels amounts to $110 \times 3 \times 10 = 3.22$ MB/s. Therefore, employing state-of-the-art CSI-based approaches with a high sampling rate or more simultaneous sensing monitors/STAs will inevitably saturate the network. In contrast, `Si-FI` ensures no redundant channel occupancy, regardless of the sampling rate or the number of sensing STAs.

6.5. *End-to-end latency analysis of `Si-FI`*

End-to-end latency is one of the most important key performance indicators for time-critical sensing applications. We analyze the end-to-end latency of `Si-FI` by breaking it down to transmission time (Tx time), inference time, and extraction time. It is important to note that the computation and transmission of BFAs are part of the mandatory channel sounding process defined in IEEE 802.11 for MU-MIMO operation. Therefore, these frames are exchanged regardless of whether `Si-FI` is deployed. `Si-FI` simply leverages the already-available feedback information for sensing, without introducing any additional transmissions, contention, or communication overhead. Consequently, the delay reported in this section purely reflects the processing latency, as the sensing framework does not alter the underlying MAC or PHY procedures. Transmission time encompasses the time taken to offload data from the sensing device to the edge server. The extraction time represents the time needed to extract a single input tensor, while the inference time denotes the time required to execute the learning block. Fig. 22 and Table 2 presents the comparative analysis of end-to-end latency of the proposed BFAs-based approach – `Si-FI`– and the `SiMWiSense` state-of-the-art CSI-based approach.

The average end-to-end latency of `Si-FI` with 234, 80, and 20 sub-channels is 30.9 ms, 22.5 ms, and 20.1 ms, respectively. This is significantly lower than the latency of the corresponding CSI-based approach, `SiMWiSense`: on average, `Si-FI` reduces end-to-end latency by 52.60 %. This substantial improvement is primarily due to the fact that, unlike `SiMWiSense`, `Si-FI` does not require any transmission, as the data is directly captured at the server. Additionally, `Si-FI` benefits from the lower extraction time of the BFAs. It is worth noting that the inference time for both the BFAs-based approach (`Si-FI`) and the CSI-based approach (`SiMWiSense`) is similar, with only about a 5 % variation. The

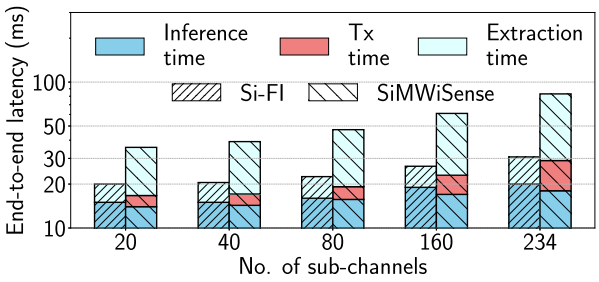


Fig. 22. `Si-FI` vs `SiMWiSense` latency comparison.

end-to-end latency for both approaches is also influenced by the tools used for BFAs and CSI extraction, as well as the computing power of the server used for extraction and inference.

7. Concluding remarks

In this paper, we propose `Si-FI`, the first framework for simultaneous multi-subject sensing using Wi-Fi beamforming feedback. Unlike CSI-based existing approaches, `Si-FI` leverages BFAs, which records the beamforming feedback information from all the simultaneous sensing STAs with a single capture at the server. In this way, `Si-FI` reduces the end-to-end latency by eliminating the need for data offloading from the sensing STAs to the server. We also introduce a BFAs-based FSL algorithm, `Si-FI FREL`, which enables fast adaptation to changes in the data distributions, enhancing the system's robustness in dynamic environments. This algorithm can generalize to new environments and subjects with only 15 seconds of new data samples, achieving an accuracy of up to 98.22 %. We evaluated the efficacy of our system using an extensive dataset collected in three different environments: a classroom, an office, and a kitchen, with three subjects performing 20 activities simultaneously. We demonstrate that `Si-FI` achieves similar performances to the CSI-based approach `SiMWiSense` on simultaneous multi-subject sensing while eliminating the channel occupation overhead and improving the end-to-end latency by 52.60 %. Additionally, with `Si-FI`, channel occupation is independent of the number of simultaneous sensing STAs and requires no firmware modifications.

Data availability

The dataset supporting the findings of this work is openly accessible at https://github.com/kfoysalhaque/Si-FI.

CRediT authorship contribution statement

Khandaker Foysal Haque: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization; **Milin Zhang:** Writing – original draft, Software, Investigation; **Francesca Meneghello:** Writing – review & editing, Validation, Supervision; **Francesco Restuccia:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

This work has been funded in part by the National Science Foundation under grants CNS-2134973, ECCS-2229472; in part by the Air Force Office of Scientific Research under contract number FA9550-23-1-0261 and in part by the Office of Naval Research under award number N00014-23-1-2221. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author(s) and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of U.S. Air Force, U.S. Navy or the U.S. Government.

References

[1] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, M. Youssef, Wideep: wifi-based accurate and robust indoor localization system using deep learning, in: 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, 2019, pp. 1–10. https://doi.org/10.1109/PERCOM.2019.8767421

[2] R. Presotto, G. Civitarese, C. Bettini, FedCLAR: federated clustering for personalized sensor-Based human activity recognition, in: 2022 IEEE International Conference on Pervasive Computing and Communications (PerCom), 2022, pp. 227–236. https://doi.org/10.1109/PerCom53586.2022.9762352

[3] B. Khaertdinov, E. Ghaleb, S. Asteriadis, Deep triplet networks with attention for sensor-based human activity recognition, in: 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), 2021, pp. 1–10. https://doi.org/10.1109/PERCOM50583.2021.9439116

[4] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, D. Zhang, Multisense: enabling multi-person respiration sensing with commodity wifi, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4 (3) (2020) 1–29.

[5] A. Khalili, A.-H. Soliman, M. Asaduzzaman, A. Griffiths, Wi-Fi sensing: applications and challenges, The J. Eng. 2020 (3) (2020) 87–97.

[6] S. Tan, Y. Ren, J. Yang, Y. Chen, Commodity wifi sensing in ten years: status, challenges, and opportunities, IEEE Internet Things J. 9 (18) (2022) 17832–17843. https://doi.org/10.1109/JIOT.2022.3164569.

[7] S.M. Hernandez, E. Bulut, wifi sensing on the edge: signal processing techniques and challenges for real-world systems, IEEE Commun. Surv. Tut. 25 (1) (2023) 46–76. https://doi.org/10.1109/COMST.2022.3209144.

[8] K.F. Haque, M. Zhang, F. Restuccia, SiMWisense: simultaneous multi-Subject activity classification through wi-Fi signals, in: 2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2023, pp. 46–55. https://doi.org/10.1109/WoWMoM57956.2023.00019

[9] A. Shokry, M. Elhamshary, M. Youssef, The tale of two localization technologies: enabling accurate low-Overhead wifi-based localization for low-End phones, in: Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2017, pp. 1–10.

[10] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, M. Youssef, Wideep: wifi-based accurate and robust indoor localization system using deep learning, in: Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2019, pp. 1–10.

[11] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, J. Han, Joint activity recognition and indoor localization with wifi fingerprints, IEEE Access 7 (2019) 80058–80068.

[12] K.F. Haque, M. Zhang, F. Meneghello, F. Restuccia, BeamSense: rethinking wireless sensing with MU-MIMO Wi-Fi beamforming feedback, Computer Networks 258, 111020 (2025), https://doi.org/10.1016/j.comnet.2024.111020.

[13] K.F. Haque, F. Meneghello, F. Restuccia, Wi-BFI: extracting the IEEE 802.11 beamforming feedback information from commercial wi-Fi devices, in: Proceedings of the 17th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, 2023, pp. 104–111.

[14] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Tool release: gathering 802.11n traces with channel state information, ACM SIGCOMM Comput. Commun. Rev. 41 (1) (2011) 53.

[15] Y. Xie, Z. Li, M. Li, Precise power delay profiling with commodity wi-Fi, in: Proceedings of the 21St Annual International Conference on Mobile Computing and Networking, 2015, pp. 53–64.

[16] F. Gringoli, M. Schulz, J. Link, M. Hollick, Free your CSI: a channel state information extraction platform for modern wi-Fi chipsets, in: Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, Association for Computing Machinery, New York, NY, USA, 2019, p. 21–28. https://doi.org/10.1145/3349623.3355477

[17] F. Gringoli, M. Cominelli, A. Blanco, J. Widmer, AX-CSI: Enabling CSI extraction on commercial 802.11ax wi-Fi platforms, in: Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & CHaracterization, 2022, pp. 46–53.

[18] Y. Tian, Y. Wang, D. Krishnan, J.B. Tenenbaum, P. Isola, Rethinking few-Shot image classification: a good embedding is all you need?, in: Proceedings of European Conference on Computer Vision (ECCV), Springer, 2020, pp. 266–282.

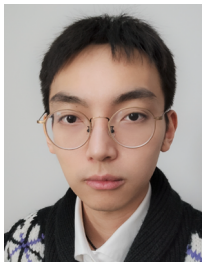
[19] J. Hu, T. Zheng, Z. Chen, H. Wang, J. Luo, MUSE-Fi: Contactless muti-person sensing exploiting near-field wi-Fi channel variation, in: Proceedings of the 29th Annual International Conference on Mobile Computing and Networking, 2023, pp. 1–15.

[20] S. Kondo, S. Itahara, K. Yamashita, K. Yamamoto, Y. Koda, T. Nishio, A. Taya, Bi-Directional beamforming feedback-Based firmware-Agnostic wifi sensing: an empirical study, IEEE Access 10 (2022) 36924–36934. https://doi.org/10.1109/ACCESS.2022.3165029

[21] T. Kanda, T. Sato, H. Awano, S. Kondo, K. Yamamoto, Respiratory rate estimation based on wifi frame capture, in: 2022 IEEE 19th Annual Consumer Communications

& Networking Conference (CCNC), 2022, pp. 881–884. https://doi.org/10.1109/CCNC49033.2022.9700721
[22] C. Wu, X. Huang, J. Huang, G. Xing, Enabling ubiquitous wifi sensing with beamforming reports, in: Proceedings of the ACM SIGCOMM 2023 Conference, 2023, pp. 20–32.
[23] Y. Ma, G. Zhou, S. Wang, H. Zhao, W. Jung, Signfi: sign language recognition using wifi, Proc. ACM on Int. Mobile, Wearable Ubiquitous Technol. 2 (1) (2018) 1–21.
[24] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, S. Guo, Wiar: a public dataset for wifi-based activity recognition, IEEE Access 7 (2019) 154935–154945.
[25] R. Xiao, J. Liu, J. Han, K. Ren, Onefi: one-Shot recognition for unseen gesture via COTS wifi, in: Proceedings of the ACM Conference on Embedded Networked Sensor Systems (SenSys), 2021, pp. 206–219.
[26] E. Yi, D. Wu, J. Xiong, F. Zhang, K. Niu, W. Li, D. Zhang, BFMSense: Wifi sensing using beamforming feedback matrix, in: 21St USENIX Symposium on Networked Systems Design and Implementation (NSDI 24), 2024, pp. 1697–1712.
[27] F. Meneghello, D. Garlisi, N. Dal 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.
[28] J. Liu, H. Liu, Y. Chen, Y. Wang, C. Wang, Wireless sensing for human activity: a survey, IEEE Commun. Surv. Tut. 22 (3) (2019) 1629–1645.
[29] Y. Ma, G. Zhou, S. Wang, Wifi sensing with channel state information: a survey, ACM Comput. Surv. (CSUR) 52 (3) (2019) 1–36.
[30] P. Ssekidde, O. Steven Eyobu, D.S. Han, T.J. Oyana, Augmented CWT features for deep learning-based indoor localization using wifi RSSI data, Appl. Sci. 11 (4) (2021) 1806.
[31] A. Dubey, P. Sood, J. Santos, D. Ma, C.-Y. Chiu, R. Murch, An enhanced approach to imaging the indoor environment using wifi RSSI measurements, IEEE Trans. Veh. Technol. 70 (9) (2021) 8415–8430.
[32] H.C. Yildirim, L. Storrer, P. De Doncker, J. Louveaux, F. Horlin, A multi-antenna super-resolution passive wi-Fi radar algorithm: combined model order selection and parameter estimation, IET Radar, Sonar Nav. 16 (8) (2022) 1376–1387.
[33] W. Li, R.J. Piechocki, K. Woodbridge, C. Tang, K. Chetty, Passive wifi radar for human sensing using a stand-alone access point, IEEE Trans. Geosci. Remote Sens. 59 (3) (2020) 1986–1998.
[34] E. Shalaby, N. ElShennawy, A. Sarhan, Utilizing deep learning models in CSI-based human activity recognition, Neural Comput. Appl. 34 (8) (2022) 5993–6010.
[35] H. Yan, Y. Zhang, Y. Wang, K. Xu, Wiact: a passive wifi-based human activity recognition system, IEEE Sens. J. 20 (1) (2019) 296–305.
[36] H. Ambalkar, X. Wang, S. Mao, Adversarial human activity recognition using wi-Fi CSI, in: 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE, 2021, pp. 1–5.
[37] M.T. Islam, S. Nirjon, Wi-Fringe: leveraging text semantics in wifi CSI-Based device-free named gesture recognition, in: 2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS), IEEE, 2020, pp. 35–42.
[38] R. Gao, M. Zhang, J. Zhang, Y. Li, E. Yi, D. Wu, L. Wang, D. Zhang, Towards position-independent sensing for gesture recognition with wi-Fi, Proc. ACM Int. Mob. Wearable Ubiquitous Technol. 5 (2) (2021) 1–28.
[39] A. Pandey, M. Zeeshan, S. Kumar, CSI-based joint location and activity monitoring for COVID-19 quarantine environments, IEEE Sens. J. 23 (2) (2023) 969–976. https://doi.org/10.1109/JSEN.2022.3196673
[40] A. Kumar, S. Singh, V. Rawal, S. Garg, A. Agrawal, S. Yadav, CNN-Based device-free health monitoring and prediction system using wifi signals, Int. J. Inf. Technol. 14 (7) (2022) 3725–3737. https://doi.org/10.1007/s41870-022-01023-7.
[41] A. Sharma, J. Li, D. Mishra, G. Batista, A. Seneviratne, Passive wifi CSI sensing based machine learning framework for COVID-Safe occupancy monitoring, in: 2021 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 2021, pp. 1–6.
[42] O.T. Ibrahim, W. Gomaa, M. Youssef, Crosscount: a deep learning system for device-free human counting using wifi, IEEE Sens. J. 19 (21) (2019) 9921–9928.
[43] S. Depatla, Y. Mostofi, Crowd counting through walls using wifi, in: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2018, pp. 1–10.
[44] J. Yang, X. Chen, D. Wang, H. Zou, C.X. Lu, S. Sun, L. Xie, Deep Learning and its Applications to WiFi Human Sensing: A Benchmark and a Tutorial, (2022). arXiv:2207.07859
[45] D. Wang, J. Yang, W. Cui, L. Xie, S. Sun, Multimodal CSI-based human activity recognition using GANs, IEEE Internet Things J. 8 (24) (2021) 17345–17355.
[46] Y. Wang, L. Yao, Y. Wang, Y. Zhang, Robust CSI-based human activity recognition with augment few shot learning, IEEE Sens. J. 21 (21) (2021) 24297–24308.
[47] S. Ding, Z. Chen, T. Zheng, J. Luo, RF-Net: a unified meta-learning framework for RF-enabled one-shot human activity recognition, in: Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 2020, pp. 517–530.
[48] Y. Jiang, X. Zhu, R. Du, Y. Lv, T.X. Han, D.X. Yang, Y. Zhang, Y. Li, Y. Gong, On the design of beamforming feedback for wi-Fi sensing, IEEE Wireless Commun. Lett. 11 (10) (2022) 2036–2040. https://doi.org/10.1109/LWC.2022.3191910
[49] E. Perahia, R. Stacey, Next Generation Wireless LANs: Throughput, Robustness, and Reliability in 802.11n, Cambridge Univ. Press, 2008.
[50] M.S. Gast, 802.11 ac: A Survival Guide: Wi-Fi at Gigabit and Beyond, O'Reilly Media, Inc., Sebastopol, CA, USA, 2013.
[51] 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. https://doi.org/10.1109/WoWMoM54355.2022.00027
[52] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al., Matching networks for one shot learning, Adv. Neural Inf. Process. Syst. 29 (2016).
[53] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, Adv. Neural Inf. Process. Syst. 30 (2017).
[54] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 1126–1135.
[55] A. Nichol, J. Achiam, J. Schulman, On First-order Meta-learning Algorithms, (2018) arXiv:1803.02999.
[56] A. Raghu, M. Raghu, S. Bengio, O. Vinyals, Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML, (2019) arXiv:1909.09157.
[57] B. Lake, R. Salakhutdinov, J. Gross, J. Tenenbaum, One shot learning of simple visual concepts, in: Proceedings of the Annual Meeting of the Cognitive Science Society, 33, 2011.
[58] S. Shi, Y. Xie, M. Li, A.X. Liu, J. Zhao, Synthesizing wider wifi bandwidth for respiration rate monitoring in dynamic environments, in: IEEE INFOCOM 2019-IEEE Conference on Computer Communications, IEEE, 2019, pp. 181–189.

Khandaker Foysal Haque is a PhD candidate in the Department of Electrical and Computer Engineering and a member of the Institute for the Wireless Internet of Things at Northeastern University, USA. He received his MS in Computer Engineering from Central Michigan University, USA, in 2021, and his BS in Electrical and Electronic Engineering from the Islamic University of Technology (IUT), Bangladesh, in 2016. His research interests include intelligent wireless systems, integrated sensing and communication, and edge computing for next-generation applications. He has published in several top-tier IEEE/ACM conferences and journals and serves as a reviewer for leading international journals and conferences. He received the Best Paper Award at IEEE iSES 2020.



Milin Zhang is a PhD student in computer engineering in the Department of Electrical and Computer Engineering and a member of the Institute for the Wireless Internet of Things at Northeastern University. He received his MS in electrical engineering from Syracuse University, USA, in 2021. He received a BS from the University of Electronic Science and Technology of China in 2018. His area of study is the integration of deep learning with emerging wireless technologies.

Francesca Meneghello (Member, IEEE) received the Ph.D. degree in Information Engineering in 2022 from the University of Padova and is currently an Assistant Professor at the Department of Information Engineering at the same university. Her research interests include deep-learning architectures and signal processing with application to remote radio frequency sensing and wireless networks. She received an honorary mention in the 2019 IEEE ComSoc Student Competition. She was a recipient of the Best Student Presentation Award at the IEEE Italy Section SSIE 2019, Best Ph.D. Thesis Award from the Italian Group of Telecommunications in 2022, and the Fulbright-Schuman Fellowship in 2023.

Francesco Restuccia (Senior Member, IEEE) is an Assistant Professor in the Department of Electrical and Computer Engineering, and a member of the Institute for the Wireless Internet of Things and the Roux Institute at Northeastern University. He received his PhD in Computer Science from Missouri University of Science and Technology in 2016, and his BS and MS in Computer Engineering with highest honors from the University of Pisa, Italy in 2009 and 2011, respectively. His research interests lie in the design and experimental evaluation of next-generation edge-assisted data-driven wireless systems. Prof. Restuccia's research is funded by several grants from the US National Science Foundation and the Department of Defense. He received the Office of Naval Research Young Investigator Award, the Air Force Office of Scientific Research Young Investigator Award and the Mario Gerla Award for Young Investigators in Computer Science, as well as best paper awards at IEEE INFOCOM and IEEE WOWMOM. Prof. Restuccia has published over 60 papers in top-tier venues in computer networking, as well as co-authoring 16 U.S. patents and three book chapters. He regularly serves as a TPC member and reviewer for several ACM and IEEE conferences and journals.