

# Device-free Crowd Sensing in Dense WiFi MIMO Networks: Channel Features and Machine Learning Tools

Sanaz Kianoush<sup>†</sup>, Stefano Savazzi<sup>†</sup>, and Monica Nicoli<sup>\*</sup>

<sup>†</sup> National Research Council of Italy

Email: [Sanaz.kianoush, stefano.savazzi]@ieit.cnr.it

<sup>\*</sup> Politecnico di Milano University

Email: Monica.Nicoli@polimi.it

**Abstract**—The paper addresses the problem of passive crowd sensing in an indoor space by processing baseband radio signals originated from a dense WiFi network. Focusing on unmodified WiFi devices equipped with multi-antenna OFDM physical radio interfaces (IEEE 802.11n/ac), we investigate the selection of statistical features to measure the body-induced alterations of the channel state information (CSI) and we analyze their dependency over the space (antennas) and frequency domains. Different machine learning methods are compared and optimized to discriminate up to 5 people moving inside the smart space. We compare different solutions for classification and target counting based on feed-forward and recurrent neural networks based on long short term memory architecture (LSTM). Experiments with real subjects are conducted to validate the proposed approach. Results confirm that CSI feature selection is crucial to optimize the counting performance and space-frequency diversity needs to be exploited to provide high-accuracy sensing in complex indoor environments

## I. INTRODUCTION

The use of built-in radio devices for device-free human body sensing is becoming attractive in many fields, such as smart living and industrial automation. Comparing with camera based solution, device-free radio sensing provides privacy-preserving monitoring with increased robustness to environmental conditions such as fire or variable light condition. In this paper, we target the problem of people counting in a confined indoor space. As depicted in the scenario of Fig. 1, detection and counting of subjects is based on real-time processing of the radio channel that is measured by a network of WiFi devices configured to transmit and receive in multiple-input-multiple-output (MIMO) configurations. Target counting is useful for retailers, space management professionals, leisure and tourism operators, passenger hubs and transport operators, to provide services for improving urban livability, workability and sustainability. For example, as transport infrastructure evolves, data from automatic passenger counting systems can help to assess the impact of any change occurred in a station or terminal, and to indicate where modifications to passenger paths are necessary to improve safety or efficiency. The use of existing wireless networks as a passive counting technology is in turn a practical way to get a clear picture of pedestrian oc-

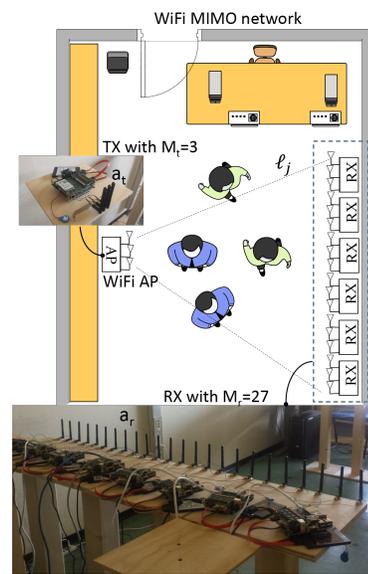


Fig. 1: Experimental network of MIMO WiFi devices. It includes 1 transmit (TX) node (serving as access point) with  $M_t = 3$  antennas and 9 receive (RX) WiFi nodes with 3 antennas each for a total number of  $M_r = 27$  receiving antennas and  $L = 3 \times 27 = 81$  sensing links.

cupancy and flow. Similarly, in retail applications, monitoring the customer traffic can help to improve the performance of stores by increasing retail conversion rates.

The proposed counting system builds upon a device-free radio sensing technology that uses receive-side measurements of the channel quality in a wireless network to detect and track any variation as possibly induced by people motion or scene alterations inside the radio coverage area. Induced fading can be detected based on the CSI estimated at the receiver side. However, learning the perturbation generated by people dynamics in complex environments is very challenging and it is thus essential to explore all the available CSI domains to maximize the sensing capability, including space (i.e., antennas), frequency and time. This motivates the employ-

which provides channel amplitude and phase information on contiguous sub-carriers of the Orthogonal Frequency Division Multiplexing (OFDM) modulation system, for every transmit-receive antenna pairs, enabling the ability to discriminate accurately the multipath characteristics.

In this paper, we consider people counting based on CSI measurements taken over multiple sub-carriers and antenna pairs (here called links) in a dense network of WiFi devices. We evaluate the performance and scalability of different machine learning (ML) methods using various CSI-related features, including common time-domain statistical features such as mean and variance [1], [2]. Numerical results show that increasing the number of people deeply affects the correlation of CSI samples observed over different antennas and/or OFDM subcarriers. We thus analyze application-specific features to effectively track the space and the frequency-domain fading correlation properties as induced by concurrently body movements. The proposed system is corroborated by experiments: we deploy a network of multi-antenna WiFi devices to monitor a whole room with the goal of counting up to 5 people inside the area. Devices collect CSI data measured at 5 GHz WiFi bands. Finally, we propose a cloud based data processing architecture that manages the CSI features collected from the dense WiFi deployment and complies with the JSON-REST web-of-things (WoT) framework. The cloud section is used to collect and classify the low-level CSI features obtained independently by the WiFi devices in real time.

The paper is organized as follows; Section II reviews the literature on target occupancy detection and counting. Section III describes the problem statement, including the MIMO radio interface and the corresponding counting problem. In Section IV, occupancy detection and counting is validated through experimental measurements. Finally, concluding remarks and open issues are summarized in Section V.

## II. RELATED WORK AND ORIGINAL CONTRIBUTIONS

Emerging technologies for radio tomographic imaging [1] and radio vision [3] rely on radio-frequency (RF) signal processing for passive (also known as *device-free*) localization [4]–[6] and activity recognition [7]. Human sensing is generally achieved by the analysis of ambient RF signals captured by densely deployed short-range wireless devices [8]. These signals are usually maintained continuously, or almost continuously, across an area of interest to enable communication tasks. Inference methods extract CSI metrics from the RF measurements to detect perturbations (due to diffraction, shadowing or scattering effects) induced by the movements of the monitored subject(s), such as people or assets [9], [10].

Similarly to imaging and localization methods, device-free crowd counting relies on the analysis of CSI data extracted at either the PHY or upper layers of a set of wireless devices inside a monitoring area [11], [12]. In [13], Wi-Fi CSI features (mean and standard deviations, or moving variance) are exploited over 30 sub-carriers for  $2 \times 2$  transmit-receive antenna pairing, along with a neural network based classifier, to count up to nine people in indoor scenarios. Compared with MAC-layer received signal strength information (RSSI), PHY-layer

CSI profiles are shown to be more effective in discriminating the number of targets in the area. Authors in [14] propose an approach to solve the key issue of the mismatching between the feature distribution during the training and testing phases, due to changing environmental conditions. They investigate feature selection considering statistics (e.g., mean, entropy), transformation based (e.g., fast Fourier transform and discrete wavelet transform) and shape based (e.g., dynamic time warping) features. Main conclusion is that redundant features not only reduce the classification accuracy but also increase the computational complexity.

In this paper, we analyze the effect of people movements and group size/dynamics on the CSI statistics obtained from a dense network of WiFi devices. In particular, we analyze the CSI over multiple antennas (27 antennas) and subcarriers (30 subcarriers). People counting, or group size estimation, is based on the analysis of CSI features, that are low dimensional representations of raw samples [15]. Differently from previous works, CSI features are here optimized considering now a MIMO-OFDM physical radio interface. Focusing on a real-time implementation, we thus consider the problem of CSI data processing over space (i.e., over multiple antennas) and frequency (i.e. across OFDM subcarriers) domains separately, and select the CSI features that are more sensitive to people motions and group size. In particular, we track the average of the CSI extra attenuation induced by the crowd movements with respect to the empty environment, in both space and frequency domains. We also introduce a new feature called antenna and sub-carrier correlation extracted from real CSI data obtained from a WiFi implementation. Finally, we propose a method based on both Feed-Forward (FF) and Recurrent Neural Networks (RNN) to count the targets using the above mentioned features. Experimental results show that an optimized design of the device-free counting methodology can reach an high accuracy (above 90%).

Best intro ever!

## III. PROBLEM STATEMENT

In this section, we discuss CSI modeling in MIMO-OFDM radio interfaces and we propose some relevant CSI-based features for occupancy inference. Data processing on CSI reports gives the opportunity to exploit *both* frequency *and* space-domain manipulation of time-series. In-fact, OFDM devices use multiple contiguous pilot sub-carriers from which CSI can be estimated at the receiver side. In addition, MIMO-enabled devices can extract CSI measurements from multiple transmit-receive antenna pairs to leverage on spatial diversity.

### A. MIMO-OFDM channel for occupancy inference

In what follows, we focus on the problem of obtaining information on the number  $C \in \{0, 1, \dots, N\}$  of human targets moving in an area  $\mathcal{X}$  served by a broadband multi-carrier MIMO network adopted for high-throughput communication. As illustrated in Fig. 1, we consider a MIMO system with an overall number of  $M_t$  transmitting antennas and  $M_r$  receiving antennas. Each target moving inside the area modifies the radio propagation between the devices and is

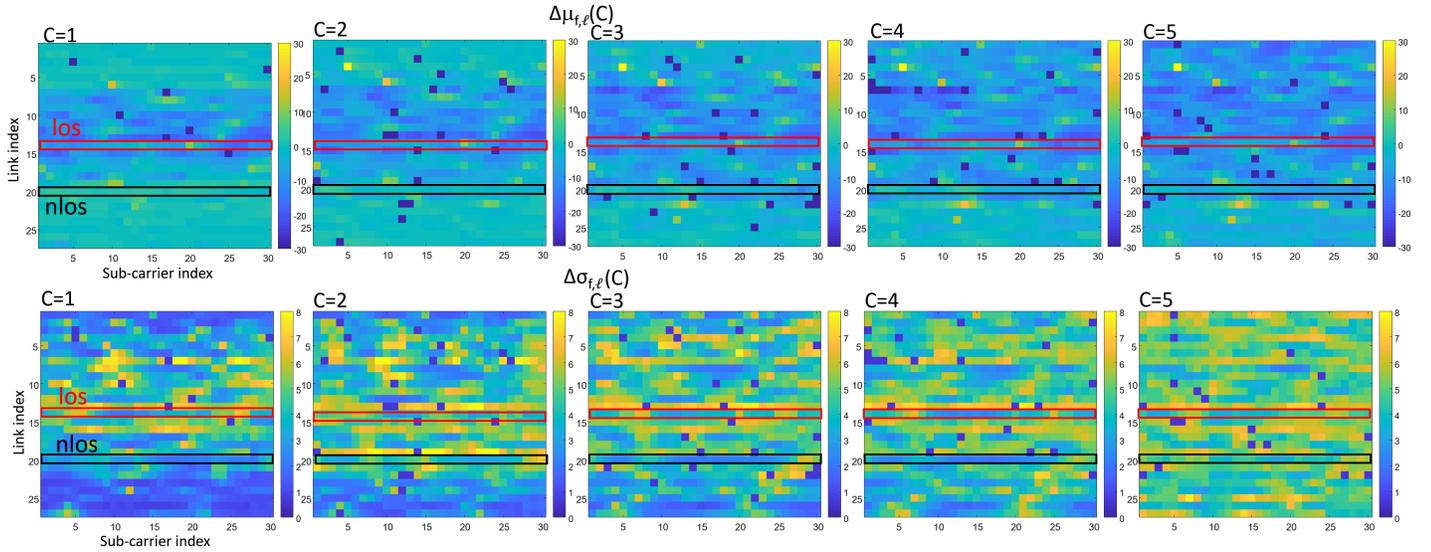


Fig. 2: Signal attenuation and perturbation profiles in the space-frequency domain for number of targets ranging from  $C = 1$  to 5. Links labeled as NLOS are obstructed by people, while LOS links are not. In the highlighted example, standard deviation for NLOS link ranges from 2 dB ( $C = 1$ ) to 7 dB ( $C = 5$ ) and attenuation from 12 to -10 dB, showing a significant sensitivity to the number of target.

expected to affect the CSI in terms of mean power, fading and also space-frequency selectivity. We thus propose to use the time-varying baseband channel response over the space-frequency domain to capture information about the presence and number of targets in the area. We assume that the MIMO channel can be estimated at the receiver devices at the discrete time instants  $t \in \mathcal{T} = \{1, 2, \dots, T\}$ , spanning over  $T$  different OFDM symbols (or frames). These observations are expected to embed a characteristic footprint of the channel variations induced by the people movement in the sensed area.

Using a conventional channel estimation method based on standard-defined pilot sub-carrier arrangement [16], the MIMO channel response is monitored in the frequency domain over  $K$  pilot sub-carriers  $f \in \mathcal{F} = \{f_1, f_2, \dots, f_K\}$ . Training/reference OFDM symbols for channel estimation are usually multiplexed with information symbols and embedded in each data frame according to the adopted WiFi standard. The complex baseband channel response observed on frequency  $f \in \mathcal{F}$ , over the link  $\ell = \ell(a_t, a_r)$  between the transmitting antenna  $a_t = 1, \dots, M_t$  and receiving antenna  $a_r = 1, \dots, M_r$ , at time  $t \in \mathcal{T}$ , is denoted as  $H_{f,\ell,t}$ , where the index  $\ell \in \mathcal{L} = \{1, 2, \dots, L\}$  ranges over the  $L = M_t M_r$  radio links. The  $M_t M_r \times 1$  space-domain channel response for the subcarrier  $f$  is  $\mathbf{H}_t^f = \{H_{f,\ell,t}\}_{\ell \in \mathcal{L}}$ , the  $K \times 1$  frequency-domain response for the link  $\ell$  is  $\mathbf{H}_t^\ell = \{H_{f,\ell,t}\}_{f \in \mathcal{F}}$ , while the overall space-frequency CSI is  $\mathbf{H}_t = \{\{H_{f,\ell,t}\}_{f \in \mathcal{F}}\}_{\ell \in \mathcal{L}}$ . CSI samples are expected to embed information about the number of targets  $C$  and are thus used in next section to extract features for target counting.

### B. Log-normal model of RSS

Features are built based on the CSI strength, or PHY-layer equivalent of the RSS, observed over the space-frequency

domain. The instantaneous channel power observed at time  $t \in \mathcal{T}$  on sub-carrier  $f \in \mathcal{F}$  and link  $\ell \in \mathcal{L}$ , is defined in dB scale as  $s_{f,\ell,t} = |H_{f,\ell,t}|_{\text{dB}}^2$ . A stochastic log-normal model is adopted to relate the perturbations of the PHY-layer RSS measurement to the target(s) number  $C$ . Considering one link and subcarrier, the instantaneous channel power can be modelled as Gaussian random variable [6]

$$s_{f,\ell,t} = \begin{cases} \mu_{f,\ell}(\Phi) + w_{f,\ell,t}(\Phi), & \text{if } C = 0 \\ \mu_{f,\ell}(C) + w_{f,\ell,t}(C), & \text{if } C \geq 1. \end{cases} \quad (1)$$

In the empty scenario case, the RSS has a deterministic mean  $\mu_{f,\ell}(\Phi) = \mu_{f,\ell}^{(p)}(\Phi) + \mu_{f,\ell}^{(m)}(\Phi)$  that accounts for path-loss  $\mu_{f,\ell}^{(p)}(\Phi)$  and static multipath effects  $\mu_{f,\ell}^{(m)}(\Phi)$ , while  $w_{f,\ell,t}(\Phi) \propto \mathcal{N}(0, \sigma_{f,\ell}^2(\Phi))$  models the random fluctuations due to measurement errors or small variations in the environment. For  $C \geq 1$ , subjects moving inside the area modify the propagation by introducing diffraction and diffusive components,  $\mu_{f,\ell}(C) = \mu_{f,\ell}^{(p)}(\Phi) - \Delta\mu_{f,\ell}^{(p)}(C) + \mu_{f,\ell}^{(m)}(C)$ , where  $\Delta\mu_{f,\ell}^{(p)}(C)$  is the attenuation caused by the target(s) and  $\mu_{f,\ell}^{(m)}(C)$  describes the body-induced multipath effects. When the targets are moving inside the sensitivity area, the received signal is also subject to an increased fluctuation due to the random term  $w_{f,\ell,t}(C) \sim \mathcal{N}(0, \sigma_{f,\ell}^2(C))$ . The target-induced attenuation  $\mu_{f,\ell}(C)$  and standard deviation  $\sigma_{f,\ell}(C)$  are

$$\mu_{f,\ell}(C) = \mu_{f,\ell}(\Phi) - \Delta\mu_{f,\ell}(C) \quad (2)$$

$$\sigma_{f,\ell}(C) = \sigma_{f,\ell}(\Phi) + \Delta\sigma_{f,\ell}(C) \quad (3)$$

where  $\Delta\mu_{f,\ell}(C) = \Delta\mu_{f,\ell}^{(p)}(C) + \mu_{f,\ell}^{(m)}(\Phi) - \mu_{f,\ell}^{(m)}(C)$  highlights the changes induced by the target presence with respect to the empty scenario, while  $\Delta\sigma_{f,\ell}(C) \geq 0$  denotes the corresponding increase of the RSS fading due to target movement. Note

that RSS observations are continuously sampled over a finite time interval (e.g., 10 ms for system validation of Sect. IV) that depends on the device duty-cycle. During this interval, targets can freely move, turn or change posture and position.

Fig. 2 shows the target-related features  $\Delta\mu_{f,\ell}(C)$  (on top) and  $\Delta\sigma_{f,\ell}(C)$  (at bottom) in the space-frequency domain for varying number of subjects ranging from  $C = 1$  to  $C = 5$ . Considering as an example the link  $\ell = 20$ , it can be seen that the RSS attenuation increases from 5 to -40 dB for  $C = 1$  to  $C = 5$ . Increasing the number of people causes incremental alterations of the propagation environment. These maps on increased RSS attenuations and standard deviations and affect the majority of the MIMO links and sub-carriers.

### C. CSI features

In this section we highlight relevant CSI features that can be used as input for counting and crowd size classification. Body movements are expected to modify the multipath propagation (e.g., the power-delay profile) and the related correlation over the space-frequency domain (e.g., the coherence bandwidth). Thereby, we consider the CSI power-frequency profile as the frequency-domain RSS vector at time  $t$

$$\mathbf{s}_{\ell,t} = [s_{f_1,\ell,t} \cdots s_{f_K,\ell,t}]^T \quad (4)$$

that collects the measurements of the channel response over all the subcarriers for link  $\ell \in \mathcal{L}$ , and, similarly, the space-domain RSS vector

$$\mathbf{s}_{f,t} = [s_{f,1,t} \cdots s_{f,L,t}]^T \quad (5)$$

collecting the channel measurements over all active MIMO links, for the pilot sub-carrier  $f \in \mathcal{F}$ . Based on log-normal modelling of individual RSS terms (1), both the frequency (4) and the space-domain (5) RSS measurements are described as multivariate Gaussian  $\mathbf{s}_{\ell,t} \sim \mathcal{N}(\boldsymbol{\mu}_\ell, \mathbf{R}_\ell)$ , while  $\boldsymbol{\mu}_\ell$  and  $\mathbf{R}_\ell$  are the mean and covariance in frequency domain and  $\mathbf{s}_{f,t} \sim \mathcal{N}(\boldsymbol{\mu}_f, \mathbf{R}_f)$ .  $\boldsymbol{\mu}_f$  and  $\mathbf{R}_f$  are the mean and covariance in space domain. Based on such framework, learning is applied to the following CSI features:

$$\begin{aligned} \boldsymbol{\mu}_\ell &= [\mu_{f_1,\ell}(C) \cdots \mu_{f_K,\ell}(C)]^T, \\ \boldsymbol{\mu}_f &= [\mu_{f,1}(C) \cdots \mu_{f,L}(C)]^T, \\ \boldsymbol{\Sigma}_\ell &= \text{diag}[\sigma_{f_1,\ell}^2(C) \cdots \sigma_{f_K,\ell}^2(C)], \\ \boldsymbol{\Sigma}_f &= \text{diag}[\sigma_{f,1}^2(C) \cdots \sigma_{f,L}^2(C)]. \end{aligned} \quad (6)$$

In addition, the corresponding matrices of CSI correlation coefficients across selected sub-carriers (frequency)  $\boldsymbol{\Gamma}_f$  and link pairs  $\boldsymbol{\Gamma}_\ell$  are also analyzed as additional CSI feature inputs for classification. These are defined as

$$\begin{aligned} \boldsymbol{\Gamma}_\ell &= \boldsymbol{\Sigma}_\ell^{-1/2} \mathbf{R}_\ell \boldsymbol{\Sigma}_\ell^{-1/2}, \\ \boldsymbol{\Gamma}_f &= \boldsymbol{\Sigma}_f^{-1/2} \mathbf{R}_f \boldsymbol{\Sigma}_f^{-1/2}. \end{aligned} \quad (7)$$

Effects of body movements on space- and frequency-domain channel correlation is still unexplored in the literature (see Sect. II). In the following sections, we deepen the topic by proving that, compared to the unobstructed environment, the presence of the target(s) affects both space- and frequency-domain channel correlation, and such alterations can be exploited as reliable features for occupancy inference.

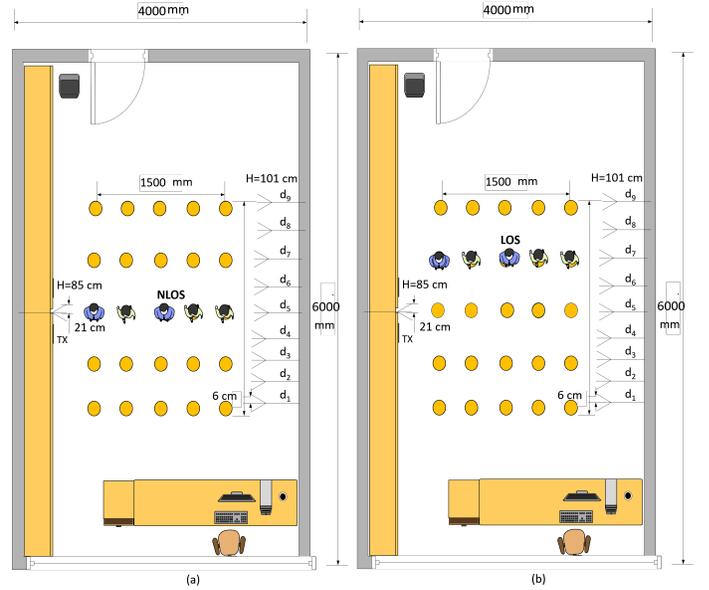


Fig. 3: Experimental layout with 1 TX and 9 RX devices. Each device has 3 antennas. The number of persons inside the area ranges from  $C=1$  to  $C=5$ . The  $C$  persons enter the area in a tight formation as in the figure, moving along a straight line from top to bottom. In LOS scenario (a), TX is not obstructed, while in NLOS scenario (b) TX is obstructed by people.

## IV. OCCUPANCY DETECTION AND COUNTING VALIDATION

The experimental activities have been conducted inside the indoor lab environment shown in Fig. 3. The network is composed of 10 WiFi devices equipped with 3 antennas each, 1 device acting as TX and 9 as RX, distributed over an area with  $6 \times 4 \text{ m}^2$  size. As shown in Fig. 3, the WiFi devices are arranged to cover the whole monitored area and to capture signals obstructed by moving objects accurately. The overall number of TX and RX antennas is  $M_t = 3$  and  $M_r = 27$ , respectively. The CSI data are collected over 30 sub-carriers. Up to  $N = 5$  people are moving inside the room as shown in Fig. 3. each record is registered The MIMO links whose line-of-sight path is obstructed by the body are referred to as non-line-of-sight (NLOS), those that are not obstructed are labeled as LOS. We employed a network of MIMO-OFDM WiFi devices configured in monitor mode and working in the 5.32GHz band (i.e., WiFi band 2, channel 64, sub-carrier spacing 312.5 kHz and nominal bandwidth equal to 20 MHz). The monitor mode allows the receiver to observe the WiFi CSI on the considered channel without explicit IP handshaking procedures. One TX device is programmed to inject (or transmit) custom IEEE 802.11n PHY protocol data units (PPDU) structured as standard high-throughput (HT) greenfield WiFi format [17] including preamble, MAC addresses, header, and payload: injected frames are sent at regular time intervals of 10 ms. In our tests, the TX device acts as access point node while the RX devices are collecting and measuring CSI reports. Modified chip-set firmware and kernel [18] have been used to obtain the CSI samples of received IEEE 802.11n data frames. The adopted chipset is the Intel

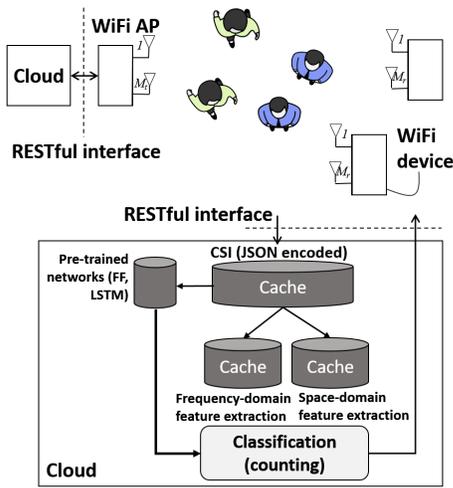


Fig. 4: Cloud platform for target counting (JSON REST interface).

Wireless Link 5300 working as a MIMO-OFDM baseband modem. The modified driver allows to extract the standard CSI reports for uni-cast/broadcast packets. As shown in the demonstration scenario depicted in Fig. 3, all WiFi devices are installed on low-power single board computers (SBC) supporting battery powered operations. Focusing on the 20 MHz bandwidth with 64 available sub-carriers, that include data, pilot and null sub-carriers, the PHY layer reports CSI data for  $K = 30$  sub-carriers, with WiFi grouping number  $N_g = 2$  and effective bandwidth equal to 18.75MHz.

The TX device multiplexes 3 spatial streams. The modulation and coding scheme (MCS) adopted for the injected frames is MCS 17, corresponding to the QPSK modulation with coding rate 1/2 [19]. Focusing on CSI estimation, a single probe transmitted from all the antennas is used at the receiver side to provide an estimate of the frequency-domain CSI terms [20] for each pilot sub-carrier,  $\hat{\mathbf{H}}_t^f = \mathbf{W}\mathbf{H}_t^f$ , multiplied by the corresponding spatial pre-coding terms  $\mathbf{W}$ . To obtain the CSI vectors  $\mathbf{H}_t^f$  to be used for occupancy inference, we post-multiply the obtained estimates  $\hat{\mathbf{H}}_t^f$  as  $\mathbf{H}_t^f = \mathbf{W}^{-1}\hat{\mathbf{H}}_t^f$ . Pre-coding  $\mathbf{W}$  is extracted from the beam-forming information and typically depends on the chip-set used [18].

In what follows, we first describe the CSI feature computing architecture, based on a REST framework (Sect. IV.A). Next we present the validation of the counting system (Sect. IV.B).

#### A. Cloud computing architecture

As depicted in Fig. 4, the deployed cloud platform exposes a set of representational state transfer (REST) application programming interfaces (APIs) that are used by the RX WiFi devices to send the CSI features encoded in a Java script object notation (JSON) format. The collection of CSI data that the cloud platform receives is thus a JSON array that contains the selected CSI features (Sect. III), as well as an indicator of frequency or space domain data structure in use. In particular, the Jax RESTful web services (Jax-RS) are adopted for exposing resources, and use the Jersey-Maven implementation. JSON parsing adds additional computation and size

		Frequency domain						Space domain						
FF		94.8 <sup>1</sup>	79.3 <sup>2</sup>	88.7 <sup>3</sup>	84.7 <sup>4</sup>	78.1 <sup>5</sup>	92.6 <sup>6</sup>	90.4 <sup>1</sup>	83.2 <sup>2</sup>	76.4 <sup>3</sup>	77.1 <sup>4</sup>	81.5 <sup>5</sup>	89.4 <sup>6</sup>	mean
		5.2	20.5	13.3	15.3	21.9	7.6	9.6	36.9	23.4	22.7	18.5	10.6	
		95.9	76.2	90.8	80.8	83.4	89.9	92.1	73.5	78.2	77.5	69.6	88.3	
		4.1	23.8	9.2	19.2	16.6	10.1	7.9	26.5	21.8	22.5	30.4	11.7	
		80.7 <sup>1</sup>	67.8 <sup>2</sup>	63.8 <sup>3</sup>	62.3 <sup>4</sup>	72.5 <sup>5</sup>	93.8 <sup>6</sup>	76.4 <sup>1</sup>	48.2 <sup>2</sup>	80.4 <sup>3</sup>	56.9 <sup>4</sup>	63.5 <sup>5</sup>	84.4 <sup>6</sup>	standard deviation
		39.3	42.2	44.2	37.7	28	6.2	23.6	51.8	39.2	43.1	36.5	15.6	
		72.2	50.4	58.4	65.9	72.2	84.1	72.3	56.8	63.2	54.8	59	82.9	
		27.8	49.6	41.6	34.1	27.8	15.9	27.7	43.2	36.8	45.2	41	17.1	
		56.6 <sup>1</sup>	90.4 <sup>2</sup>	37.8 <sup>3</sup>	38.1 <sup>4</sup>	43.5 <sup>5</sup>	100 <sup>6</sup>	97.7 <sup>1</sup>	96.7 <sup>2</sup>	97.3 <sup>3</sup>	97.3 <sup>4</sup>	98.5 <sup>5</sup>	98.4 <sup>6</sup>	correlation
		43.4	69.6	62.2	71.9	57	0	2.3	3.3	2.8	2.2	2	1.2	
		50.9	41.9	36.5	34.5	32.7	92.8	97.7	96.8	97.3	97	98.2	99.2	
		49.1	58.1	63.5	65.5	67.3	7.2	2.3	3.2	2.7	3	1.8	0.8	

(a)

		Frequency domain						Space domain						
LSTM		98.1	97.2	99.3	100.4	97.5	99.6	98.1	99.2	100.3	99.4	97.5	100.6	mean
		2	3	1	0	3	1	2	1	0	12	3	0	
		100	97	96.1	97.1	100	100	100	99	98	96.7	89	100	
		0	3	3.9	3.9	0	0	0	1	2	3.3	11	0	
		99.1	88.2	88.3	35.4	35.5	84.6	99.1	96.2	94.3	92.4	90.5	92.6	standard deviation
		1	12	32	65	45	16	1	4	6	8	10	8	
		100	88.9	82.9	44.9	49.5	64.1	100	96	9.3	92.9	89.1	93.9	
		0	1.1	17.1	55.1	50.5	35.9	0	4	8.7	7.1	10.9	6.1	

(b)

Recall, False negative rate, Precision, False positive rate

Fig. 5: Target counting results in terms of recall, precision, false positive and false negative rates, using FF-NN (a) and LSTM (b) methods, based on CSI features extracted over the frequency (left) or space (right) domains, including mean, standard deviation and correlation. The identified 6 classes correspond to the empty environment  $C = 0$  (class labelled as '1') or crowd size of  $C = 1, 2, 3, 4, 5$  people (class labels '2', '3', '4', '5', '6').

overhead as the result of object serialization/deserialization stages, however many open source libraries can be found to optimize the performance [22].

Once the cloud platform has received the CSI features, it unpacks and sends them to a database which is used by the target counting methods as described in Sect. IV.B. Finally, the cloud platform exposes some application programming interfaces (APIs) to query the system and to obtain aggregated CSI data for debugging.

#### B. Counting results and validation

In this section, experimental results are presented according to the layout described in Sect IV. For the purpose of training the crowd sensing system, we collected 600 CSI samples over a time window of 10 seconds, for each antenna pair and each of the  $K = 30$  sub-carriers. Two neural network (NN) methods are applied and compared for people counting inside the monitored indoor space using as inputs the features described in Sect III-B. In particular, we compared Feed-Forward (FF-NN) and Recurrent NN (RNN) approaches. The FF-NN is a practical solution that can be used for conventional classification problems. The input layer takes features defined in (6), namely the frequency-domain  $(\mu_\ell, \Sigma_\ell)$  or the space-domain  $(\mu_f, \Sigma_f)$  means and deviations, or the corresponding correlation coefficients  $(\Gamma_\ell, \Gamma_f)$  in (7). Input layer is followed by a single hidden layer with 10 neurons, a softmax layer and

an output classification layer with 6 classes, for counting up to  $N = 5$  people moving in the monitored area (including unoccupied room). The RNN is implemented using an LSTM architecture. This is specifically designed to model temporal sequences and, differently from hidden Markov models, includes long-range dependencies. LSTM architectures have been explored for large-scale acoustic modeling in speech recognition, language translation, and handwriting recognition [23]. The methodology is here used to track long and short-term dependencies on temporal sequences of CSI features (i.e., over space and frequency domains). Based on the features defined in (6), the LSTM input layer extracts temporal sequences by applying time averaging sequentially over a moving window of  $T = 100$  samples.

For both FF-NN and LSTM, 80% of data is used for training the network, 10% for validation and 10% for testing the classifiers. Fig. 5 shows the classification results in terms of recall, false negative rate (FNR), precision and false positive rates (FPR) [5] based on frequency and space domain inputs for the FF-NN (Fig. 5(a)) and LSTM (Fig. 5(b)). In particular, focusing on the detection of the empty environment by FF-NN in Fig. 5(a), the recall results for the method using mean, standard deviation and correlation features are 95%, 61% and 57% in frequency domain and 90%, 76% and 97% in space domain, respectively. Tracking of antenna correlations (over space domain) gives the best result for the considered settings.

Notice that, the computation of the correlation coefficients of all sub-carrier and antenna pairs would increase the complexity compared with mean and deviation features. We thus apply an optimization policy to select an optimal group of sub-carriers and antenna (links) pairs for which correlations are extracted and processed. In particular, the considered antenna pairs are  $\ell(a_t = 1, a_r = 1)$ ,  $\ell(a_t = 2, a_r = 2)$   $\ell(a_t = 2, a_r = 14)$ , sub-carriers are  $f_5, f_{15}$  and  $f_{30}$ .

Considering LSTM, data processing automatically extracts long to short-term time-domain correlations in the input features: correlation features (7) are thus not used as network inputs. As shown in Fig. 5 (b), LSTM improves recall value to about 99% in frequency domain and 98% in space domain respectively. Also, LSTM gives more reliable results in terms of recall, FNR, precision and FPR. As far as counting problem is considered, the results highlight that space diversity provides higher accuracy. Moreover, correlation information is the most reliable feature for target counting using FF-NN while LSTM has better performance in terms of recall, FPR, FNR and precision values with respect to FF-NN.

## V. CONCLUSION

This paper proposes CSI feature selection over the sub-carriers and multiple links of dense indoor WiFi MIMO-OFDM networks, along with machine learning methodologies, for device-free crowd sensing applications, namely for people counting. Statistical features including mean, standard deviation and correlation coefficients, are extracted from channel observations taken over the space-frequency domain, and they are used as input for FF-NN and LSTM learning methods. Experimental tests have been carried out using a WiFi MIMO

infrastructure to evaluate the performance of the proposed methods. A cloud-based computing platform has been developed to manage the aggregation of the CSI features from the dense WiFi network. Future work will focus on integration with high-frequency technologies for monitoring of larger population sizes.

## REFERENCES

- [1] J. Wilson et al., "Radio tomographic imaging with wireless networks," *IEEE Trans. on Mobile Computing*, vol. 9, no. 5, pp. 621-632, 2010.
- [2] S. Depatla et al., "Occupancy estimation using only WiFi power measurements," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1381-1393, Jul. 2015.
- [3] S. Savazzi et al., "Device-free radio vision for assisted living: Leveraging wireless channel quality information for human sensing," *IEEE SP Mag.*, vol. 33, no. 2, 2016.
- [4] M. Bocca et al., "Multiple target tracking with RF sensor networks," *IEEE Trans. on Mobile Computing*, vol. 13, no. 8, pp. 1787-800, Aug. 2014.
- [5] Kianoush et al., "Device-Free RF human body fall detection and localization in industrial Workplaces," *IEEE IoT J.* vol. 4, no. 2, pp. 351-362, April 2017.
- [6] S. Savazzi et al., "A Bayesian approach to device-free localization: Modeling and experimental assessment," *IEEE Journal on Selected Topics in Signal Processing*, vol. 8, no. 1, pp.16-29, Feb. 2014.
- [7] S. Sigg et al., "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. on Mobile Computing*, vol. 13, no. 4, pp. 907-920, Apr. 2014.
- [8] B. Guo et al., "Behavior recognition based on Wi-Fi CSI: Part 1 and 2," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 90-90, Oct. 2017.
- [9] G. Kouitatis, "Multiple human effects in body area networks," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 938-941, 2010.
- [10] V. Rampa et al., "EM models for passive body occupancy inference", *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 16, pp. 2517-2520, 2017.
- [11] Iker Sobron et al., "Device-free people counting in IoT environments: New insights, results and open challenges," *IEEE Internet of Things Journal*, DoI:10.1109/JIOT.2018.2806990, Feb, 2018.
- [12] S. Bartoletti et al., "Device-Free counting via wideband signals," *IEEE J. sel. areas Comm.*, vol. 35, no. 5, 2017.
- [13] Yen-Kai et al., "Device-free indoor people counting using Wi-Fi channel state information for Internet of things," *IEEE Global Communications Conference (GLOBECOM)* 2017.
- [14] H. Zou et al., "FreeCount: Device-free crowd counting with commodity WiFi," *IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6, 2017.
- [15] S. Kianoush et al., "A cloud-IoT platform for passive radio sensing: Challenges and application case studies," *IEEE Internet of Things Journal*, DoI:10.1109/JIOT.2018.2834530, May, 2018.
- [16] M. Cicerone et al., "Channel estimation for MIMO-OFDM systems by modal analysis/filtering," *IEEE Transactions on Communications*. vol. 54, no. 11, pp. 2062-2074, Nov. 2006.
- [17] H. A. Omar et al., "A Survey on high efficiency wireless local area networks: Next generation WiFi," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2315-2344, 2016.
- [18] D. Halperin et al., "Tool release: Gathering 802.11n traces with channel state information." *ACM SIGCOMM CCR* 41, Nr. 1, S. 53, 2011.
- [19] IEEE Std 802.11n-2009, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE, New York, NY, USA, Oct 2009.
- [20] R. Crepaldi et al., "CSI-SF: Estimating wireless channel state using CSI sampling & fusion," *Proc. of IEEE INFOCOM*, pp. 154-162, Orlando, FL, Mar 2012, doi: 10.1109/INFOCOM.2012.6195523
- [21] Di Domenico et al., "Trained-once device-free crowd counting and occupancy estimation using wifi: A Doppler spectrum based approach," *IEEE WiMob*, 2016.
- [22] K. Maeda, "Performance evaluation of object serialization libraries in XML, JSON and binary formats," *Proc. of Digital Information and Communication Technology and it's Applications (DICTAP)*, pp. 177-182, 2012.
- [23] R. Zazo et al., "Age Estimation in Short Speech Utterances Based on LSTM Recurrent Neural Networks," *IEEE Access* vo.6, pp. 22524-22530, 2018