Human Presence and Activity Detection from WiFi CSI Data using CNN

Md Touhiduzzaman
Department of Computer Science
Virginia Commonwealth University
Richmond, Virginia, USA
touhiduzzamm@vcu.edu

Maher Al Islam
Dept. of Electrical & Computer Engineering
Virginia Commonwealth University
Richmond, Virginia, USA
alislamm@vcu.edu

Samah Ahmed
Department of Computer Science
Virginia Commonwealth University
Richmond, Virginia, USA
ahmedss5@vcu.edu

***Abstract*—WiFi sensing can be used to monitor specific human movements and also to predict the movements after learning it using any machine learning paradigm. In this project, a deep neural network is implemented for a pruned large dataset of human mobility through a WiFi network delivering CSI data. The hyperparameter optimization technique is used to find out specific values of several parameters of the machine learning model to detect human presence and several movement activities with XX% of accuracy.**

***Index Terms*—DNN, WiFi, sensing, CSI, machine learning, human activity detection, surveillance**

I. Introduction

In this era of conflict between surveillance and privacy concerns, WiFi sensing brings a very optimal lineup to monitor necessary surveillance data pruning through unnecessary personal details by the implementation. It also benefits through lower data storage and processing time than camera recorded video or image data. Therefore, to implement any surveillance system where specific movements are needed, one need not record heavy cam-feeds and occupy huge processing overhead, rather can use WiFi channel state information (CSI) data to learn the specification of the movements and then classify those.

Channel State Information (CSI) obtained from commercial WiFi chipsets has proven to be efficient in detecting human interactions inside any radio wave. Due to high availability & practical usability, the WiFi network can be the most effective radio network to collect CSI data. The popularity of approaches that measure Received Signal Strength (RSS) by narrowband radio devices is due to cost-effectiveness and ubiquity. However, recent progress in signal descriptors, such as CSI, obtained through low-cost chipsets like ESP32, Intel 5300 NIC, and IEEE 802.11n chipsets offer enhanced accuracy compared to RSS. Due to high temporal variance in RSS, slow movements of humans end up hidden in the inherent signal variability [4]. Comparatively, the structure of CSI is temporally more stable than RSS because it captures small-scale multipath propagation over multiple sub-carriers in frequency domain [5] [6]. CSI indicates different physical qualities of the channel, such as shadowing, frequency selective fading, multipath propagation, and interference effects. Hence, CSI is currently a good alternative for RSS. Compared to traditional technologies, CSI-based detection has several advantages. It detects humans through walls, does not depend on lighting, preserves user privacy, and importantly, occupants are not required to carry any devices. Hence, it is widely used to quantify the human presence interaction with the wireless channel in the form of occupancy detection, activity/gesture/identity recognition, and human positioning.

The attributes of WiFi CSI data vary depending on routers and network specifications. The values are also hugely impacted by the orientation of the environment, i.e., the room structure, furniture in the room, openings as doors or windows in the room, moving items like flying curtains, ambient or any other sound, etc. These variations are difficult to characterize by naked human realization. That is why a blind machine learning paradigm like a deep neural network, convolutional neural network, and so on can be used to train and classify the CSI dataset. In this project, a deep neural network (DNN) with an optimized number of hidden layers is implemented. The neural network trains a weight array by the larger portion of the dataset and then implies the weight array directly over the test dataset, which is a subset of the whole collection of the CSI dataset.

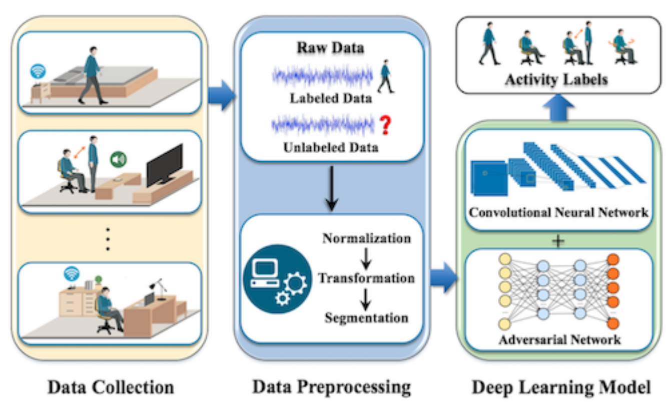

Fig. 1. System Framework

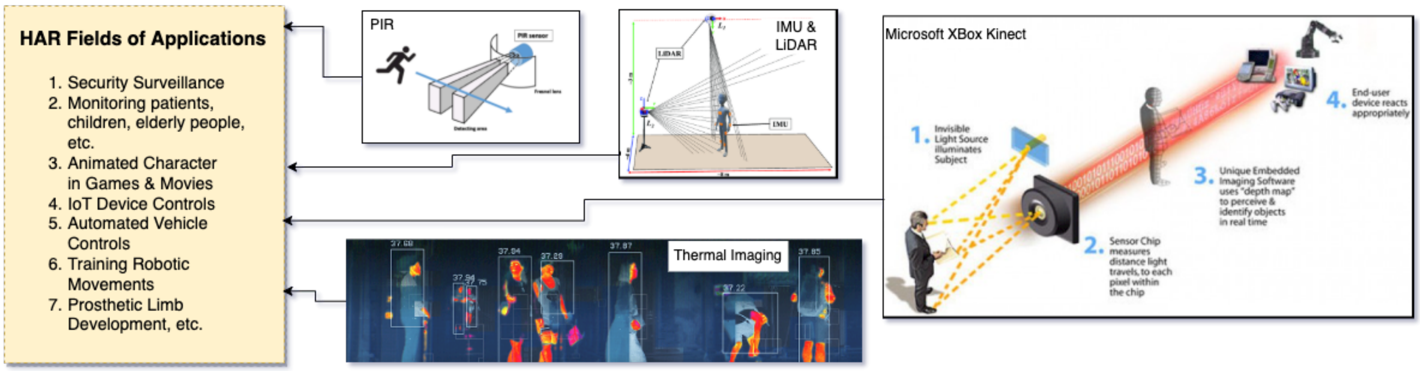

Fig. 2. Different Applications of HAR using Several Sensors (Image Credits [11] [18])

II. Background & Related Works

Human Activity Recognition (HAR) has been a very much explored research arena. These researches have been spanned from surveillance & monitoring to animated movie and gaming, controlling Internet of Things (IoT) devices through various gestures, automated vehicle controls, robotic evolution, etc. Most of the researches have been done using camera feeds, smart wearable devices and using several other sensors like - infrared (IR), passive infrared (PIR), Microsoft XBOX Kinect, IMU (Inertial Measurement Unit) [8], thermal imaging, ultrasound, pyroelectric infrared [9] [10], etc. However, there are only a few researches done using WiFi sensed CSI data. All of these researches can be discussed in three sections -

1) Various methods of HAR,
2) HAR using WiFi sensing CSI data and
3) Novelty of our work

A. *Various Methods of Human Activity Recognition*

Various custom sensors are experimented to detect several human activities in many studies and several are already in the public market, like Microsoft XBOX Kinect, IMU, PIR, Rokoko smartwears [16], etc.

Even for a very subtle movements like fingers & wrists, we can refer some quite ingenious studies using custom made sensors. In [12], a mmWave based radio frequency is shown to be analyzable and so learnable using CNN that rehab movements of any injured one of the 19 joints in human hand can be monitored using the custom made wireless sensor. An comparison with radar imaging is also shown in that study, where their implementation is claimed to be performing better than that in this specific case of hand rehab. This sensor is also claimed to be consuming least amount of energy ($443\mu J$) and least amount of time ($64\mu s$) than any sensors present in the market.

In another similar study [13], a sensor is designed using mathematical derivation for least Root Mean Square Error (RMSE) to detect one wide-arm exercise of any human. The advantage of the sensor is claimed to be of miniature in size and very much portable. In another study [14], a wearable soft smargloves is designed as a sensory & motor glove. It can classify 16 separate finger gestures used in mirror therapy & task-oriented therapy with upto 93.32% accuracy using SVM, kNN & DT algorithms.

Another study uses Multi-Dimensional Dynamic Time Warping (MDTW) model using MatLAB model and Leap Motion (LM) [17] sensors to detect four hand gestures using a robotic hand simulation. Another study [15] uses IMU sensors to classify several movements of shoulder, hip, knee and elbow. Dynamic Time Warping (DTW) is used in this study.

Moving towards bigger movement detection, we can refer to Dalal et. al. [19], where any human movement in a motion capture video recording is shown to be detectable for even examples having more than 4400 human. Yin et. al. [20] uses only 2 pairs of Pyroelectric Infrared sensors to classify the speed, distance and direction any human movement with 94% accuracy. In another research by Park et. al. [21], it is shown by experiments that impulse radar signals can be used to detect human presence and motion with almost 100% accuracy. Jeong et. al. [22] developed a probabilistic method to detect human subjects using even a low-resolution thermal imaging sensor. It covers various pre and post-image processing methods including background collection, Gaussian filtering, segmentation, local/global adaptive threshold and background learning. Patil et. al. [18] uses a combination of multiple LiDAR sensors Inertial Motion Unit (IMU) sensors to identify various poses for a human in motion, even at real time. States of the different moving joints of human limbs for different walking style are classified using multiple IMU sensors in another study by Semwal et. al. [23]. In general, these used sensors have been very costly or designed using a complex process that may be difficult to maintain as per the course of usage. All of these studies can be countered by our proposal with only one statement that - our proposed WiFi sensing setup can be extended to function without any kind of extra hardware setup, but only by using the traditional WiFi routers, which can be easily found around us now a days. Therefore, our proposed system may outperform these studies in the matter of practical usability.

B. *Human Activity Recognition using WiFi Sensing*

There have been comparatively very few studies using WiFi sensing as a medium of detecting human activities and gestures. Wu & Zhang et. al. proposes DeMan [24] system, which is a unified scheme for non-invasive detection of moving and stationary human on commodity WiFi devices. Extensive experimental evaluation in typical indoor environments validates the performance of DeMan in various human poses and locations and diverse channel conditions. Particularly, DeMan provides a detection rate of around 95% for both moving and stationary people, while identifies human-free scenarios by 96%. But this study skips any specific action of a moving human being inside their proposed network.

Domenico et. al. [25] proposes a WiFi-based through-the-wall (TTW) presence detection of stationary and moving humans by analyzing the doppler spectrum. However, their proposal includes two WiFi routers on each side of the wall, which may be impractical in many scenarios. Another study by Hernandez et. al. [26] proposes a one-sided wall occupancy monitoring, where it is shown that a single WiFi signal receiver device is enough to sense any human walking even on the opposite side of any wall and keep a count of them. In another study by Hernandez et. al. [27], a system called WiFedarate is proposed, which implements an edge based federated learning over WiFi sensing. The proposed WiFederated system can scale as more locations and scenarios for learning are added and it provides a more accurate and time efficient solution compared to existing transfer learning and adversarial learning solutions using the parallel training ability at multiple clients. By introducing new client selection methods during the federated-learning process, the accuracy is shown to increase further and then the feasibility of training models at the edge and introduce continuous annotation to allow for continuous learning over time is evaluated to be good.

WiFi network's CSI data is explored in a study by Palipana et. al. [28]. Using CSI data, they have modeled the human presence and then analysed the occurrences of non-linear correlations among the WiFi sub-carriers. Then they have exploited these correlations by introducing non-linear techniques to reduce CSI dimensions and filter noise, which can detect human presence and movements even when human motion is insignificant or occurs very far from the link. Their non-linear techniques improve the detection accuracy up to 5% compared to the linear approach with just two transceivers.

A very interesting study by Liu et. al. [29] proposes a CSI analyzing technique which can detect static human through estimating the breathing frequency by exploring phase information of CSI. They achieved more robust data by fusing subcarriers and filtered out environmental noise by adopting Butterworth filter and using hampel filter before and during wavelet denoising. For estimating the frequency, they introduced Fast Fourier Transformation (FFT), Estimating Signal Parameter via Rotational Invariance Techniques (ESPRIT) and Multiple Signal Classification (MUSIC). The results of their study show that human detection accuracy can achieve higher than 95% and averaged evaluating accuracy can reach 89.8% with their system.

However, none of these studies have covered the detection of human presence and movements in a specific room with any neural network based learning.

C. *Novelty of Our Work*

As we have already have discussed in II-A, the proposed sensors are either costly, consume extra power, complex to design and maintain, costly to maintain or costly to process the information. So, those custom sensory developments can be trumped over by our approach with more commonly available and maintainable WiFi networks.

Moreover, none of the human detection systems using WiFi sensing, as discussed in II-B, considers both stationary and moving human at the same time and processes it with more complex signal processing algorithms. Our proposal excludes any complex preprocessing of the raw CSI data, other than calculating the plain amplitude of the signals from those data, and then runs a deep neural network model over them to detect any human presence and movement. From this point of view, our work is novel and different from any other existing approaches to detect human motion.

III. Problem Statement

There are various implementations of human activity recognition. However, the purpose of this problem is to identify as many common human actions as can be detected, like - entering into a room, walking around the room, standing or sitting idle in the room, exiting the room, etc. In other words, we can state the problem as - if a person P does any action a_i, such that $a_i \in A = \{a_1, a_2, a_3, ..., a_n\}$ within a specific environment E, the problem requires us to classify that action within a predefined set of classes $C = \{c_1, c_2, c_3, ..., c_n\}$.

IV. Solution Proposal

We have divided the problem stated in III into three subproblems as shown in Figure 1, like -

1) Data Collection
2) Data Preprocessing
3) Deep Learning Model

Our solution apporaches for all of these subproblems are described in the following subsections.

A. *Data Collection*

In our specific implementation, we have concentrated on implementing an efficient deep learning model, i.e., the third subproblem, while a pre-processed proven dataset is used from an open-source collection [7]. During the dataset collection process as shown in [7], ten activities are done in four specific room setups by two different people. Therefore, referring to section III, we can specify our problem variables as -

1) Set of persons, $P = \{p_1, p_2\}$,
2) Set of environments, $E = \{e_1, e_2, e_3, e_4\}$,
3) Actions, $A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\}$
4) Set of classes, $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\}$

The schematics of the four rooms are shown in Figure 3 as specified by $G.21$, $G.19$, $1.28A$ & 2.60. Inside each of those rooms, an Intel 5300 NIC Device [3] acts as the access-point (AP), while three Raspberry Pi clients connect with the AP. CSI & RSS data are collected for each of those three client-AP connection pairs. We have not used RSS data as it is shown CSI dataset is enough to classify actions rather than complicating things using RSS [4] [5] [6]. The CSI dataset produced by Intel 5300 NIC devices has 30 subcarriers and produces one complex number per transmitter-receiver pair, i.e., there are 30 complex numbers per CSI data in our dataset. These data need to be processed using some interpolation techniques and denoised using any denoising filter like DWT or Hampel, which is discussed in the following subsection.

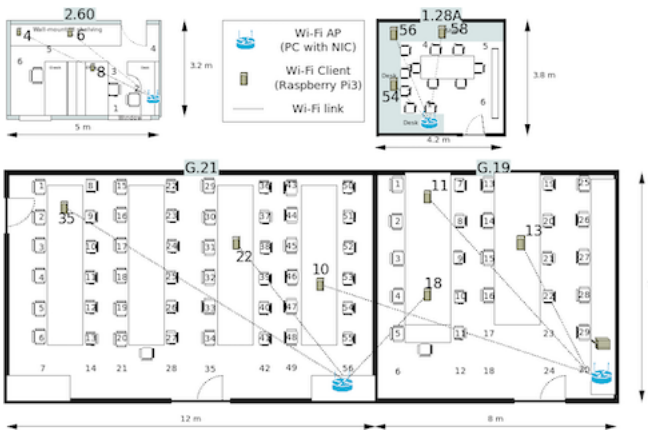

Fig. 3. Schematics of the Experiment Rooms

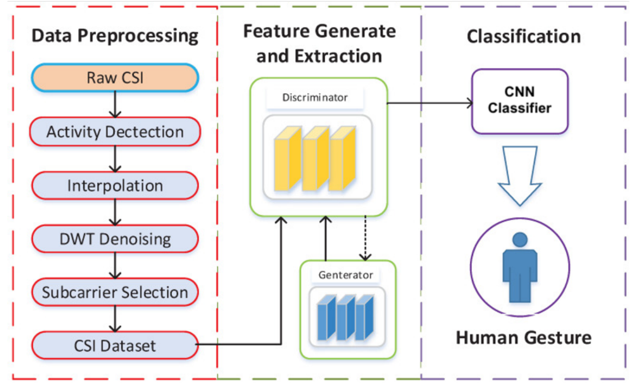

Fig. 4. Steps to Classify WiFi Sensed CSI data for Human Actions

B. Data Preprocessing

The CSI data provided by the Intel 5300 NIC is a set of complex numbers, which need to be processed by interpolating amplitude, phase and frequencies. Over that resultant interpolation, we can apply any denoising filter to smooth the wave forms. We have selected Discrete Wavelet Transform (DWT) filter for this case. Afterward, we need to select specific number of subcarriers ranging from 1 to 30 and select varying number of CSI data per batch to feed into our classifier. In between these preprocessing and classification steps, we can have an extra step to generate and extract specific features out of the filtered CSI data as amplitude, frequency, phase, etc. Another step before feeding these data to the classifier would be to split it into 70% training set & 30% test set.

C. Deep Learning Model

Once the CSI data are denoised, we can make a machine learn from the dataset using any neural-network algorithm. We choose Convolutional Neural Network (CNN) to run over the dataset and classify each action among any of the ten predefined classes of human gestures. As shown in the steps the figure 5, we need to run multiple iterations of the model evaluation until a predefined number of epochs with the same dataset. The resultant model would be tested again by modifying any of the parameters of the model as well as of the dataset like number of subcarriers and CSI windows size. At a certain point of these optimization steps, we are most likely to have a model delivering high accuracy during evaluation of the model using our test dataset. At that point, we can finalize the model and complete our experiment.

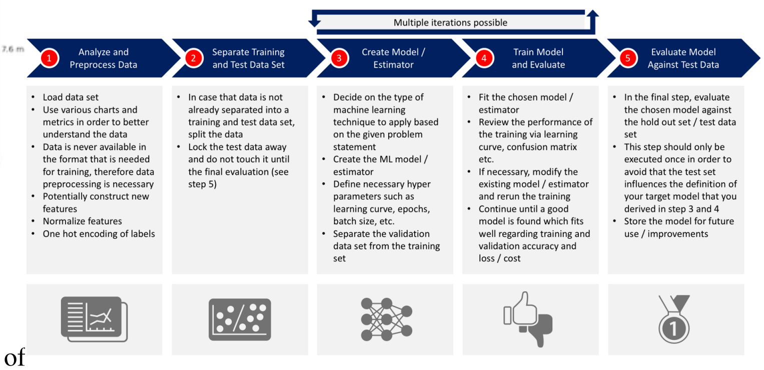

Fig. 5. Data Preprocessing to ML Training & Testing Steps [11] [18]

REFERENCES

[1] Jiang W, Miao C, Ma F, Yao S, Wang Y, Yuan Y, Xue H, Song C, Ma X, Koutsonikolas D, Xu W. Towards environment independent device free human activity recognition. InProceedings of the 24th Annual International Conference on Mobile Computing and Networking 2018 Oct 15 (pp. 289-304).

[2] Palipana, Sameera, Piyush Agrawal, and Dirk Pesch. "Channel state information based human presence detection using non-linear techniques." Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments. 2016.

[3] https://dhalperi.github.io/linux-80211n-csitool/

[4] Zheng Yang et al. "From rssi to csi: Indoor localization via channel response." *ACM Comput. Surv.*, 46(2):25:1–25:32, December 2013.

[5] Z. Zhou et al. Towards omnidirectional passive human detection. In Proc. of INFOCOM, 2013.

[6] Jiang Xiao et al. Find: Fine-grained device-free motion detection. In Proc. of ICPADS, 2012.

[7] Anonymous Author. (2020). A Wi-Fi Channel State Information (CSI) and Received Signal Strength (RSS) data-set for human presence and movement detection (0.2.2) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3677366

[8] Semwal, V.B., Gaud, N., Lalwani, P. et al. Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor. Artif Intell Rev 55, 1149–1169 (2022). https://doi.org/10.1007/s10462-021-09979-x

[9] Q. Hao, D. J. Brady, B. D. Guenther, J. B. Burchett, M. Shankar and S. Feller, "Human Tracking With Wireless Distributed Pyroelectric Sensors," in IEEE Sensors Journal, vol. 6, no. 6, pp. 1683-1696, Dec. 2006, doi: 10.1109/JSEN.2006.884562.

[10] Q. Hao, F. Hu and Y. Xiao, "Multiple Human Tracking and Identification With Wireless Distributed Pyroelectric Sensor Systems," in IEEE Systems Journal, vol. 3, no. 4, pp. 428-439, Dec. 2009, doi: 10.1109/JSYST.2009.2035734.

[11] https://iesedmonton.org/news/2018-5-22/what-are-pir-sensors-and-why-do-i-need-them, https://www.wired.com/2010/11/tonights-release-xbox-kinect-how-does-it-work/, https://www.sourcesecurity.com/tags/thermal-imaging.html

[12] Sizhe An and Umit Y. Ogras. 2021. MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare. ACM Trans. Embed. Comput. Syst. 20, 5s, Article 72 (October 2021), 22 pages. DOI:https://doi.org/10.1145/3477003

[13] R. Abbasi-Kesbi and A. Nikfarjam, "A Miniature Sensor System for Precise Hand Position Monitoring," in IEEE Sensors Journal, vol. 18, no. 6, pp. 2577-2584, 15 March15, 2018, doi: 10.1109/JSEN.2018.2795751.

[14] X. Chen et al., "A Wearable Hand Rehabilitation System With Soft Gloves," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 943-952, Feb. 2021, doi: 10.1109/TII.2020.3010369.

[15] Ana Pereira, Vânia Guimarães, and Inês Sousa. 2017. Joint angles tracking for rehabilitation at home using inertial sensors: a feasibility study. In ¡i¿Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare¡/i¿ (¡i¿PervasiveHealth '17¡/i¿). Association for Computing Machinery, New York, NY, USA, 146–154. DOI:https://doi.org/10.1145/3154862.3154888

[16] https://www.rokoko.com/

[17] https://www.ultraleap.com/

[18] Patil, Ashok Kumar & Balasubramanyam, Adithya & Ryu, Jae & B N, Pavan Kumar & Chakravarthi, Bharatesh & Chai, Young-Ho. (2020). Fusion of Multiple Lidars and Inertial Sensors for the Real-Time Pose Tracking of Human Motion. Sensors (Basel, Switzerland). 20. 10.3390/s20185342.

[19] Dalal N., Triggs B., Schmid C. (2006) Human Detection Using Oriented Histograms of Flow and Appearance. In: Leonardis A., Bischof H., Pinz A. (eds) Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science, vol 3952. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11744047_33

[20] Yun, J.; Lee, S.-S. Human Movement Detection and Identification Using Pyroelectric Infrared Sensors. Sensors 2014, 14, 8057-8081. https://doi.org/10.3390/s140508057

[21] Young-Jin Park, Hui-Sup Cho (2020) A Method for Detecting Human Presence and Movement Using Impulse Radar, Division of Electronics & Information System, DGIST, Daegu, 42988, South Korea

[22] Y. Jeong, K. Yoon and K. Joung, "Probabilistic method to determine human subjects for low-resolution thermal imaging sensor," 2014 IEEE Sensors Applications Symposium (SAS), 2014, pp. 97-102, doi: 10.1109/SAS.2014.6798925.

[23] Semwal, V.B., Gaud, N., Lalwani, P. et al. Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor. Artif Intell Rev 55, 1149–1169 (2022). https://doi.org/10.1007/s10462-021-09979-x

[24] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu and J. Cao, "Non-Invasive Detection of Moving and Stationary Human With WiFi," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 11, pp. 2329-2342, Nov. 2015, doi: 10.1109/JSAC.2015.2430294.

[25] S. Di Domenico, M. De Sanctis, E. Cianca and M. Ruggieri, "WiFi-based through-the-wall presence detection of stationary and moving humans analyzing the doppler spectrum," in IEEE Aerospace and Electronic Systems Magazine, vol. 33, no. 5-6, pp. 14-19, May-June 2018, doi: 10.1109/MAES.2018.170124.

[26] S. M. Hernandez and E. Bulut, "Adversarial Occupancy Monitoring using One-Sided Through-Wall WiFi Sensing," ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500267.

[27] S. M. Hernandez and E. Bulut, "WiFederated: Scalable WiFi Sensing using Edge Based Federated Learning," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3137793.

[28] Sameera Palipana, Piyush Agrawal, and Dirk Pesch. 2016. Channel State Information Based Human Presence Detection using Non-linear Techniques. In ¡i¿Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments¡/i¿ (¡i¿BuildSys '16¡/i¿). Association for Computing Machinery, New York, NY, USA, 177–186. DOI:https://doi.org/10.1145/2993422.2993579

[29] Y. Liu, H. Liu and F. Gao, "Channel State Information-based Device-Free stationary Human Detection with estimating respiratory frequency," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2019, pp. 1111-1117, doi: 10.1109/IAEAC47372.2019.8997815.