Comparative Study of Person Identification Using Sub-6 GHz and mmWave Wi-Fi CSI

Maksim Karnaukh
Dept. of Computer Science
University of Antwerp
Antwerp, Belgium
maksim.karnaukh@student.uantwerpen.be

***Abstract*—Person identification plays a vital role in enabling intelligent, personalized, and secure human-computer interaction. Recent research has demonstrated the feasibility of leveraging Wi-Fi signals for passive person identification using a person's unique gait pattern. Although most existing work focuses on sub-6 GHz (5 GHz) frequencies, the emergence of mmWave (60 GHz) Wi-Fi offers new opportunities through its finer spatial resolution (5mm wavelength vs. 6cm at 5GHz), though its comparative advantages for person identification remain unexplored. This work presents the first comparative study between sub-6 GHz and mmWave Wi-Fi signals for person identification with commodity hardware, using a custom-collected dataset of synchronized measurements from both frequency bands in an indoor environment. To ensure a fair comparison, we apply identical training pipelines and model configurations across both frequency bands. By doing experiments with multiple deep learning architectures in an end-to-end manner, we demonstrate that for low-rate sampling (10 Hz), 60 GHz signals can achieve superior performance (91.2% average accuracy for 20 individuals) when proper background subtraction is applied, particularly for larger groups of individuals. However, high-rate 5 GHz sampling (200 Hz) can very well match low-rate mmWave performance in certain configurations. We study the trade-offs between frequency bands in terms of data efficiency, temporal resolution requirements, and robustness to environmental variations.**

***Index Terms*—Person Identification, millimeter-wave, Sub-6 GHz, Wi-Fi Sensing, Gait Recognition, CSI, Deep Learning, End-to-End Learning**

I. Introduction

In recent years, Wi-Fi-based human identification has emerged as an interesting and rapidly evolving area of research, driven by the increasing availability of Wi-Fi-enabled devices and the growing demand for non-intrusive, passive authentication methods in smart environments. Traditional biometric systems, such as fingerprint [1], iris [2], and facial recognition are quite accurate, but they often require dedicated hardware, active user participation, or raise privacy concerns. Wearable device-based [3], [4] and vision-based [5], [6] identification methods on the other hand offer a slight alternative, but they also have limitations: wearables require user compliance, and vision systems are constrained by lighting conditions, line-of-sight requirements, and again privacy issues.

To address these limitations, researchers have begun using the ubiquitous presence of Wi-Fi signals in (indoor) environments. As people move through these spaces, their bodies subtly alter the radio frequency (RF) signals, through reflection, scattering, and absorption, in a way that is unique to their physical body build and walking style [7]–[9]. These effects are captured through fine-grained Channel State Information (CSI), which describes how wireless signals propagate across different subcarriers and antenna paths, and which can be collected manually [10]–[13]. Human gait, in particular, has proven to be a promising biometric here, namely, it involves full-body movement, is difficult to impersonate, and does not need any special user interaction or additional devices [14]–[16].

Building on these insights, systems such as WiFi-ID [17], WiWho [9], and WifiU [18] demonstrated early success in identifying individuals by analyzing gait cycles and RF perturbations. However, many of these relied on hand-crafted features (and also shallow classifiers), limiting their scalability and robustness. More recent approaches, such as NeuralWave [19], HumanFi [20], and Gate-ID [21], have tried deep learning to automatically learn discriminative gait features from raw CSI data. NeuralWave introduced a deep CNN architecture, HumanFi employed LSTM-based temporal modeling, and Gate-ID incorporated attention mechanisms to handle directional gait patterns and antenna placement effects, boosting classification accuracy.

Although most of this research has focused on sub-6 GHz Wi-Fi, there is a growing interest in exploring the potential of mmWave Wi-Fi for human sensing and identification tasks. mmWave offers higher spatial resolution and finer signal granularity due to its shorter wavelength, which may enable a more precise capture of motion-induced variations in wireless signals. Some early efforts have used radar-based mmWave systems to perform multi-person detection and identification [22]. Others have investigated the feasibility of using Commercial Off-The-Shelf (COTS) IEEE 802.11ad devices for gesture recognition [23]. However, the trade-offs between COTS mmWave and sub-6 GHz systems for person identification, particularly in terms of robustness and generalization, seem to remain underexplored to our knowledge.

In this paper, our aim is to address this by conducting a comparative study of person identification performance using both traditional sub-6 GHz and mmWave Wi-Fi signals. Using end-to-end deep learning techniques, without the need for manual feature extraction, we analyze CSI data from both frequency bands to evaluate how well each system

can distinguish people based on their gait. To enable this comparison, we collected a unique synchronized multi-modal CSI dataset using COTS hardware, with 5 GHz and 60 GHz measurements recorded simultaneously from 20 participants in an indoor setting. By systematically comparing the two, we investigate whether the higher resolution of mmWave offers measurable benefits for person identification, and we note the potential limitations or challenges.

This paper makes the following key contributions:

- **Feasibility Study on mmWave CSI:** We evaluate the effectiveness of mmWave Wi-Fi signals for person identification using a private dataset of 7 individuals collected with COTS 60 GHz devices.
- **Multi-modal CSI Dataset Collection:** We collect a new 'multi-modal' CSI dataset containing synchronized 5 GHz (sub-6 GHz) and 60 GHz (mmWave) measurements in a shared environment, enabling direct comparison between the two frequency bands for the human identification task.
- **Cross-band Performance Comparison:** We present the first experimental comparison of mmWave and sub-6 GHz Wi-Fi CSI for person identification. Using several deep learning architectures in an end-to-end manner, we evaluate their accuracy under identical training conditions to better understand the trade-offs between the two technologies.
- **Benchmarking with Deep Models:** We implement and benchmark multiple temporal learning architectures including Long Short-Term Memory networks (LSTMs), residual Convolutional Neural Networks (CNNs), and Temporal Convolutional Networks (TCNs). These models are inspired in part by previous work on sub-6 GHz Wi-Fi sensing and adapted for mmWave CSI to explore their effectiveness.

The remainder of this paper is organized as follows. Section II describes the problem statement and our central research question. Section III reviews prior work on sub-6 GHz and mmWave-based sensing. A basic technical overview is provided in Section IV. Section V describes our data collection process and experimental setup, together with the used models and preprocessing techniques. Section VI presents a comprehensive evaluation across multiple frequency bands and model types. We discuss some of our limitations and problems in Section VII, and finally, we conclude the paper and outline future directions in Section VIII.

II. Problem Statement and Research Question

The core problem addressed in this paper is: *How can mmWave Wi-Fi signals be used to improve person identification, and how do they compare to traditional 5 GHz signals in terms of performance and robustness?*

This leads to our central research question:

> **To what extent can mmWave Wi-Fi outperform sub-6 GHz Wi-Fi in person identification tasks based on wireless sensing of gait?**

III. Related Work

The emergence of fine-grained CSI from commodity Wi-Fi devices has enabled a wave of research into RF-based sensing. By capturing how wireless signals interact with the environment, including reflections, diffractions, and absorptions caused by human movement, Wi-Fi has proven to be a powerful medium for passive sensing in indoor settings.

A. CSI-based Wireless Sensing

Building on this foundation, research has explored a wide range of applications that leverage CSI for passive sensing. One prominent area is activity recognition, where systems have demonstrated the ability to detect basic motions [24]–[26] and activities such as sitting, walking, or falling [27]–[30]. Further advances have enabled gesture recognition, allowing users to interact with devices through subtle hand movements without the need for physical contact [31], [32].

Beyond physical (big movement-based) sensing, CSI has also been used in sensitive health-related applications, such as monitoring respiration and heart rate in a contactless manner [33], [34]. Researchers have even demonstrated the ability to infer fine motor activities, including keystrokes on a keyboard or smartphone [35].

Next to activity and gesture recognition, CSI-based sensing has also found applications in indoor localization [36], [37], enabling device-free tracking and positioning of individuals within a space. All of these applications showcase the versatility of Wi-Fi sensing in various domains, and although these systems have shown promising results, their focus has primarily been on detecting actions or behaviors, rather than distinguishing between individuals themselves. Person identification poses a possibly differently nuanced challenge, requiring the extraction of fine-grained, person-specific features from the RF signals.

B. Gait-based Person Identification

Recent work has explored the potential of using wireless signals for biometric identification based on gait, the unique walking style of each individual. These systems are built on the idea that the interaction between a walking person and the surrounding RF field produces a distinctive signature that can be captured via CSI.

1) Sub-6 GHz Approaches: WiFi-ID [17] was among the first to show that variations in CSI caused by walking could serve as a biometric signature. By analyzing changes in the amplitude and phase of the signal, the system could distinguish between individuals. However, it relied on handcrafted features and traditional classifiers, which limited scalability.

WiWho [9] improved upon this by segmenting walking sequences into step cycles and extracting step and walk features as part of their step and walk analysis module. These features were then used for classification. Notably, WiWho demonstrated that reliable identification could be achieved even when subjects walked a few steps, reporting an average accuracy of 92% and 75% for 2 and 7 people, respectively.

WifiU [18] took a slightly different approach, i.e., it generated spectrograms from CSI measurements to resemble Doppler radar outputs and extracted gait features from the CSI spectrograms.

Despite their promising results, these systems generally relied on hand-engineered features and shallow classification methods, possibly limiting their robustness to changes in environment, group size, and walking conditions. To address these limitations, more recent research has shifted toward deep learning approaches that automatically extract discriminative features from 'raw' CSI data.

NeuralWave [19] introduced a 23-layer convolutional neural network to learn robust latent representations of gait patterns directly from preprocessed CSI samples. This approach significantly improved scalability and accuracy, achieving 87.76% $\pm$ 2.14% identification accuracy on a dataset of 24 users.

HumanFi [20] adopted a temporal modeling strategy using a six-layer LSTM-based neural network. The model improved accuracy by fusing both amplitude and phase information from CSI.

Gate-ID [21] addressed practical deployment challenges such as walking direction and antenna orientation. It proposed an attention-based architecture that incorporated both residual CNNs and bidirectional LSTMs. By simulating mirrored walking directions through data augmentation and modeling spatial–temporal dynamics, Gate-ID achieved identification accuracies between 90.7% and 75.7% for group sizes ranging from 6 to 20 people, respectively.

These deep learning-based systems demonstrate significant improvements in robustness, generalizability, and scalability, paving the way for more practical applications of Wi-Fi-based gait recognition in real-world smart environments.

2) mmWave Approaches: There has also been growing interest in leveraging the unique properties of mmWave signals for human sensing tasks. Due to their shorter wavelengths and higher spatial resolution, mmWave systems are capable of capturing more detailed movement-induced variations in wireless signals, which can certainly be advantageous for distinguishing between individuals.

Most of the work in this domain currently relies on mmWave radar. For instance, Gu et al. proposed mmSense [22], a device-free system that uses 60 GHz mmWave radar to detect and identify multiple people based on environmental fingerprints and human body reflections, achieving person identification using LSTM-based models by analyzing outline profiles and vital signs. Similarly, MGait [38] employs a 77 GHz Frequency Modulated Continuous Wave (FMCW) radar to extract gait features from micro-Doppler signatures, enabling identification and intruder detection in multi-person environments with up to 88.6% accuracy for 5 subjects.

Other radar-based systems have combined multimodal data to improve recognition performance. An interesting example is the Cross Vision-RF Gait Re-identification framework [39], which fuses mmWave radar with RGB-D camera inputs to overcome modality gaps and achieve 92.5% top-1 accuracy on a dataset of 56 participants. Another study [40] introduced a multimodal fusion algorithm that integrates mmWave radar-derived features such as phase, respiration, and heartbeat signals with an attention-based deep network, achieving 94.26% accuracy on a dataset of 20 participants.

Despite these advances in radar-based systems, research using mmWave Wi-Fi (e.g., IEEE 802.11ad or 802.11ay) for person identification remains limited. One of the few works that explores mmWave Wi-Fi sensing is by Bhat et al. [23], which investigates the use of 60 GHz beam Signal-to-Noise Ratios (SNRs) as sensing features for gesture and pose recognition. The authors compare the performance of these beam SNRs with CSI collected from a 5 GHz Wi-Fi access point, using a deep neural network to classify gestures relevant to extended reality (XR) applications. While their results demonstrate strong classification accuracy (up to 96.7% in a single environment), the work focuses on gesture recognition and does not extend to the task of person identification. Another relevant study, though focused on localization, is by Blanco et al. [41], who demonstrate the viability of using mmWave CSI from commercial MikroTik wAP 60Gx3 devices. Their system, MultiLoc, leverages both sub-6 GHz and 60 GHz CSI to enhance localization accuracy. Although MultiLoc targets localization, their work also highlights the potential of COTS mmWave Wi-Fi equipment for the task of person identification.

In summary, although mmWave systems, particularly radar-based, have shown strong potential for person identification via gait, the use of mmWave Wi-Fi CSI for it remains an underexplored area. Research using mmWave CSI, though for other sensing domains, remains important for us.

IV. Background and Technical Overview

A. Channel State Information (CSI)

In wireless communications, signals propagate from a transmitter to a receiver through complex paths involving reflection, scattering, diffraction, and absorption. These multipath effects are formally captured by the Channel State Information (CSI), which describes how a wireless channel distorts the signal over time and frequency.

CSI is typically extracted using known pilot symbols embedded between data in Wi-Fi packets, enabling receivers to estimate the channel response. In Orthogonal Frequency-Division Multiplexing (OFDM)-based systems, such as IEEE 802.11ac/ad, CSI provides the channel frequency response across multiple subcarriers and antenna paths. Modern Wi-Fi systems often employ Multiple-Input Multiple-Output (MIMO) technology, allowing CSI to be measured across multiple transmit and receive antennas.

Mathematically, the wireless channel can be modeled as:

$$Y = H \cdot X + N \tag{1}$$

where Y is the received signal, X is the transmitted signal, H is the complex CSI matrix, and N represents noise.

For a single subcarrier k, the CSI matrix $H(k) \in \mathbb{C}^{N_r \times N_t}$ captures the complex channel gains between N_t transmit and N_r receive antennas. Each entry $h_{rc}(k)$ in it, representing the

gain from the c-th transmitting antenna to the r-th receiving antenna, can be decomposed into:

$$h_{rc}(k) = |h_{rc}(k)|e^{j\theta_{rc}(k)} \tag{2}$$

where $|h_{rc}(k)|$ denotes the amplitude and $\theta_{rc}(k)$ the phase shift introduced by the channel. [42]–[44]

Thus, CSI for a packet can be viewed as a 3D matrix of dimensions $N_r \times N_t \times K$, with K being the number of subcarriers. Over time, a sequence of such matrices forms a 4D tensor, which can be analyzed for temporal patterns caused by human motion.

In our work, we focus on the amplitude component $|h_{rc}(k)|$. Although phase information can offer additional detail, it requires careful calibration. Prior research has shown that amplitude alone contains sufficient discriminative information for tasks such as person identification using deep learning.

B. CSI at Sub-6 GHz vs. mmWave Frequencies

Most existing research on Wi-Fi-based sensing and person identification has relied on sub-6 GHz CSI, particularly at 2.4 GHz and 5 GHz bands. These frequencies offer long-range communication and strong wall penetration, making them robust in cluttered (non-line-of-sight) environments. However, due to their relatively long wavelengths (12.5 cm at 2.4 GHz and 6 cm at 5 GHz), they exhibit coarse spatial resolution, which may limit the sensing granularity needed to capture fine movement features like gait compared to mmWave.

In contrast, mmWave Wi-Fi, operating at 60 GHz (wavelength of $\sim$5 mm), offers significantly finer spatial resolution and wider bandwidth (up to 2.16 GHz), enabling the system to capture subtle motion-induced signal variations. This would make 60 GHz Wi-Fi theoretically ideal for biometric sensing tasks such as gait recognition. However, mmWave suffers among others from shorter range and strict(er) line-of-sight requirements.

We should note that mmWave communication is currently significantly less widespread in both commercial and residential deployments compared to sub-6 GHz Wi-Fi. Given that sub-6 GHz GHz is already present in most Wi-Fi-enabled devices, systems based on sub-6 GHz CSI have a clear advantage in terms of accessibility and deployability. Although our work explores the benefits of mmWave for person identification, we acknowledge the limited adoption of 60 GHz Wi-Fi in real-world applications.

C. Deep Learning for CSI-Based Sensing

Deep learning has become a dominant approach for learning discriminative patterns in complex data domains such as images, audio, and time series. In the context of wireless sensing, deep neural networks, particularly 1D Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have shown strong performance in modeling temporal dynamics and spatial dependencies within CSI data.

1D CNNs are effective at capturing local temporal features in CSI sequences. These networks apply convolution operations over the time dimension, allowing them to learn translation-invariant features. Given an input signal x and a kernel w of size k, the output of a 1D convolution at time t, with stride 1 and appropriate padding, is computed as:

$$y(t) = \sum_{i=0}^{k-1} w(i) \cdot x(t + i - \lfloor k/2 \rfloor) \tag{3}$$

This operation extracts local temporal patterns centered around time step t, and includes both past and future context, depending on the kernel size and padding strategy, which is termed a non-causal operation. [45], [46]

In contrast, a causal convolution ensures that the output at time t depends only on the current and past inputs. This is used in architectures like Temporal Convolutional Networks (TCNs).

LSTMs, on the other hand, are a type of recurrent neural network (RNN) well-suited for learning long-term temporal dependencies. Each LSTM cell maintains a memory state through a combination of input, forget, and output gates. A simplified formulation of the LSTM cell is:

$$\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}$$

where x_t is the input at time t, h_t is the hidden state, c_t is the cell state, and σ denotes the sigmoid activation function. [45]

More recently, attention mechanisms [47] have been introduced to further improve model performance by allowing the network to selectively focus on the most informative parts of a sequence. This often leads to better performance in tasks involving noisy data.

These deep learning architectures form the foundation of some of the state-of-the-art methods for wireless sensing tasks like person identification.

V. System Design and Methodology

In this work, we evaluate person identification capabilities using two datasets containing CSI from Wi-Fi signals: (1) a multi-modal dataset we collected ourselves using both 5 GHz (sub-6 GHz) and 60 GHz (mmWave) devices, and (2) an external (private) mmWave-only dataset. This approach enables us to not only validate our methods on our controlled environment but also to first assess (person identification) feasibility and performance on an external, independent dataset.

A. External mmWave Dataset

First, to evaluate the feasibility of mmWave-based person identification on independent hardware and environment, we use a private dataset, as mentioned above, not developed by us. This dataset was collected using MikroTik wAP 60G devices equipped with 60 GHz antenna arrays.

Each device contains a Uniform Rectangular Array (URA) of 6×6 antenna elements, though only 30 out of the 36 are usable due to hardware limitations. CSI extraction was performed using an open-source toolchain. The CSI data consists of amplitude-only values for the 30 antenna elements per frame. The part of the dataset we will use includes walking trajectories from 7 different people and background samples recorded when no person was present in the environment.

The effective sampling rate in this dataset is 22 packets per second.

B. Collected Dataset (5 GHz and 60 GHz)

To evaluate and compare the person identification capabilities of sub-6 GHz and mmWave Wi-Fi signals, we collected a multi-modal CSI dataset using both 5 GHz and 60 GHz hardware. Our experimental setup was designed to simultaneously capture signal variations from both frequency bands as participants walked through the designated sensing area.

1) Physical Setup: Our data collection took place in an indoor environment over the course of three days, completing data collection from 20 participants: 7 on day one, another 7 on day two, and the remaining 6 on day three. Each participant walked continuously for two minutes in a linear path, turning around at each end and walking back, creating a round-trip pattern through the sensing area, as shown in Fig. 1(b). We also collected background signal samples, where no person was present in the environment, on the third day only. Fig. 1(a) shows the exact distances between the devices and such.

For the sub-6 GHz setup, we deployed two ASUS RT-AC86U routers. One acted as an Access Point (AP), while the other was configured as a passive CSI monitor. An Intel laptop transmitted Internet Control Message Protocol (ICMP) echo requests at approximately 300 packets per second to the AP, which responded with echo replies. The monitor router, placed opposite the AP, captured the CSI from these responses. This setup, similar to the setup in [48] and illustrated in Fig. 2, recorded CSI measurements in the 5 GHz band using a single transmit and single receive antenna configuration. To enable a fair comparison with the mmWave system, the packet transmission rate was reduced to 10 packets per second after processing, matching the sampling frequency of the 60 GHz devices.

For the mmWave component, we used two pairs of 60 GHz Wi-Fi devices (wAP 60G) placed diagonally across from each other, forming an "X" (cross-shaped layout). Each device pair operated independently, capturing packets at a rate of approximately 10 packets per second.

2) Participants: Concerning the participants, a total of 20 volunteers were asked for the data collection process. The group consisted of people between the ages of 20 and 30. The height, weight, and sex distributions of the participants are shown in Fig. 3.

3) CSI Extraction and Firmware Modifications: For 5 GHz CSI extraction, the ASUS RT-AC86U routers were modified using the Nexmon CSI toolchain[1], which is adapted from [49], [50]. Nexmon is a firmware patching framework that enables CSI access on Broadcom-based chipsets used in commodity Wi-Fi routers. Specifically, we used the Nexmon CSI extension developed for the RT-AC86U model, which provides access to per-packet CSI data for 802.11ac transmissions over 80 MHz bandwidth.

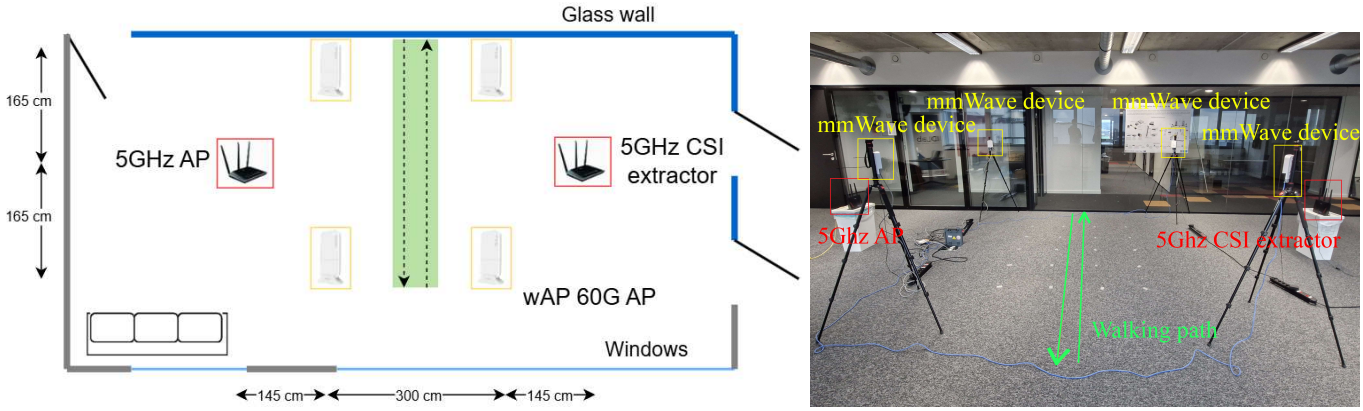
We used the Nexmon CSI firmware version 10_10_122_20 patched for the Broadcom BCM4366c0 chip. The patching process involved building custom firmware using the OpenWRT build environment and flashing the routers accordingly. After flashing, the routers were configured to operate in monitor mode using Nexmon's kernel modules, allowing one router to passively capture CSI data from the ICMP echo replies sent between the Intel laptop and the AP. We apply Medium Access Control (MAC) filtering to capture the packets from the expected device. All our experiments were conducted using an 80 MHz channel in the 5 GHz band, fixed to channel 36. The exact steps executed can be found in our project code [2]. This setup gave us CSI measurements across 64 subcarriers, of which 52 are data subcarriers. The remaining 12 subcarriers, consisting mainly of pilot and null subcarriers, are excluded from analysis during postprocessing, as they do not carry useful information.

For the 60 GHz setup, CSI-like measurements were obtained using a set of custom mmWave access points running a modified OpenWRT firmware developed by IMDEA Networks [51]. Data collection was done using custom scripts [52] enabling synchronized packet transmission and logging. Each mmWave link captured approximately 10 packets per second. It is important to note here that our mmWave data sampling rate is constrained by a device-specific firmware limitation, not by an inherent property of the frequency band. Each device pair produced an array of 30 useful amplitude values per time step, resulting in a total of 60 amplitude values across the two active pairs.

Environmental conditions were kept as consistent as possible across the three days of data collection to reduce variation due to external interference or background motion. All collected data was timestamped and stored locally for processing, i.e. synchronization, alignment, etc. To ensure temporal alignment between the 5 GHz and 60 GHz streams, we first parsed and converted the CSI files from the mmWave frequency band into structured CSV format with ISO-formatted timestamps. The 60 GHz data was collected in parallel from two devices and merged by matching packets with timestamps within a one-second window. The 5 GHz PCAP files were decoded into CSI amplitude matrices using the Nexmon CSIKit tool [11], and a delta timestamp correction (extracted from file metadata) was applied to account for clock offsets. We then synchronized the two frequencies using their overlapping timestamp ranges and adjusted for any systematic timing shifts (full-hour differences due to device desynchronization). Finally, we downsampled both data streams to 10 Hz and linearly interpolated

[1]https://github.com/nexmonster/nexmon_csi

[2]https://github.com/MaksimKarnaukh/MasterThesis_mmWavePI/blob/master/csi_collection/readme.md



(a) (b)

Fig. 1: (a) Floor plan and walking area and (b) real scene settings.

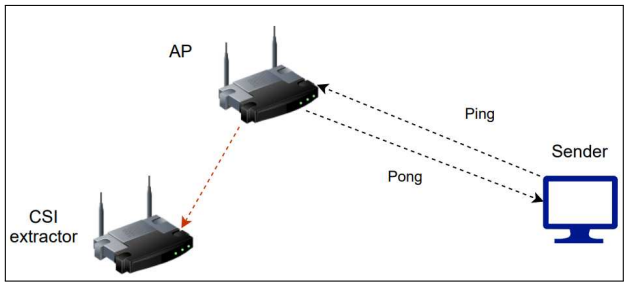


Fig. 2: Experimental setup for the sub-6 GHz CSI data collection. The devices are positioned to capture motion along a linear walking path. Figure adapted from [48].

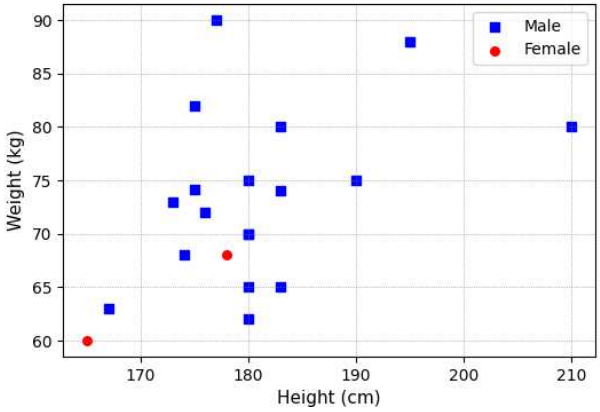


Fig. 3: Participant statistics (N=20, ages 20-30) showing height, weight and sex distributions.

occasional missing values caused by rare measurement failures (resulting in infinite-amplitude subcarrier values). These instances were sparse and isolated, and interpolation was applied to ensure consistency in each amplitude array of the time series, before saving the synchronized amplitude-only CSI matrices as NumPy arrays for our model training. Interpolation should not have a significant effect on our results because of the small scale and nature of OFDM systems.

C. Data Preprocessing and Augmentation

We adopt an end-to-end learning approach, minimizing the amount of manual feature engineering or domain-specific transformation. The raw CSI signals are directly fed into deep neural networks after minimal preprocessing. The only preprocessing operation performed is **background subtraction** [53], where an average CSI sequence captured in the absence of any person, as we collected, is subtracted from the walking data. This emphasizes dynamic variations caused primarily by human motion and tries to suppress static environmental artifacts.

To improve generalization and mitigate overfitting, we use the following data augmentation techniques during training:

1) Gaussian Smoothing: To simulate natural signal variation and also possibly reduce high-frequency noise, we apply Gaussian smoothing as a form of data augmentation. Specifically, a one-dimensional Gaussian filter is convolved over the time axis of the CSI sequence, slightly blurring the signal in time. This encourages the model to focus on more robust, lower-frequency motion patterns. The smoothing is applied probabilistically during training, and the filter's kernel size and standard deviation are treated as hyperparameters.

2) Mixup: We also incorporate the Mixup augmentation technique, proposed by Zhang et al. [54], which generates synthetic training examples by convexly combining pairs of input signals and their corresponding labels. Given two samples (x_i, y_i) and (x_j, y_j), a new sample is generated as:

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j, \quad \tilde{y} = \lambda y_i + (1-\lambda)y_j \tag{4}$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$. We will use $\alpha = 0.4$ as a default value. This augmentation encourages the model to behave linearly in-between training examples and improves robustness against noise and overfitting. Mixup is meant to be applied only during training, and the combined pairs are chosen randomly in each batch.

When both Gaussian smoothing and Mixup are enabled for a particular configuration, we apply them in sequence: Gaussian smoothing is applied first to each input sample, followed by Mixup. This order ensures that smoothing acts on realistic, temporally coherent signals, while Mixup then blends those smoothed signals across examples. Applying Mixup first could interfere with the interpretability of the interpolated motion patterns if gaussian smoothing were applied afterward.

D. Model Architectures

To assess the feasibility of mmWave CSI-based person identification and to explore temporal modeling strategies, we evaluate several deep learning architectures spanning both recurrent and convolutional paradigms. As a starting point, we draw some inspiration from models previously proposed in the context of sub-6 GHz CSI-based person identification, adopting and extending them to fit the mmWave domain for our comparative purposes. This ensures that we have experimental baselines that are grounded in architectures already shown to be effective in similar tasks. All models are trained from scratch using supervised learning and receive sequences of amplitude-only CSI samples as input, allowing for end-to-end learning without handcrafted features. We adopt a 70%-15%-15% split for training, validation, and testing, respectively.

1) LSTM and Variants: We implement a Long Short-Term Memory (LSTM) model following the architecture proposed in the HumanFi paper [20], where a unidirectional LSTM processes the input signal and the hidden state of the final timestep is passed to a dense classifier. We extend this baseline in several ways:

- **Bidirectional LSTM (BiLSTM)** to capture both past and future context.
- **CNN-BiLSTM with Attention**, which augments the BiLSTM with a CNN layer in front and attention mechanisms to selectively weight the temporal features. Variants include temporal attention and dual attention. The latter is based on the Zhu et al. [55] implementation.

2) Residual CNN: To exploit hierarchical temporal features in the CSI sequences, we look at the residual convolutional neural network (ResNet) inspired by the architecture used in [56], originally proposed for 'joint activity recognition and indoor localization' (JARIL). The base model consists of 4 layers, each composed of one residual block (i.e., a (1,1,1,1) configuration). Each block follows the classic two-layer bottleneck design with batch normalization and ReLU activation, and downsampling is applied at the beginning of each layer when needed to adjust the resolution or channel depth.

We explore several variants of this ResNet-style architecture:

- **OptResNet1D-JARIL:** A refined version of the original JARIL ResNet. We retain the (1,1,1,1) configuration, but we replace the aggressive initial downsampling with a more conservative stem (stride 1) to preserve fine-grained temporal details. The model includes four residual layers with progressively doubled channel widths, and the model ends with global average pooling and a fully connected output layer.
- **OptECAResNet1D-JARIL:** An extension of OptResNet1D-JARIL, enhanced with Efficient Channel Attention (ECA) [57] in each residual block. The ECA modules allow the network to adaptively reweight feature channels based on their temporal relevance, improving robustness to noise and focusing attention on discriminative gait features.
- **Custom ResNet:** A simplified variant with a consistent stride across all layers, reducing complexity and improving computational efficiency. We also evaluate an ECA-augmented version of this model.

3) Temporal Convolutional Network (TCN): We also experiment with a Temporal Convolutional Network [58], a causal 1D CNN with dilated convolutions and residual blocks. TCNs are known to outperform recurrent models on sequence modeling tasks due to their parallelism and ability to model long-range dependencies efficiently. We adopt the generic configuration evaluated in [58] with an added classification head and adjust the number of levels, dilation rates, and kernel sizes.

VI. Experimental Results

In this section, we present and analyze the results of our person identification models across both datasets and multiple experimental setups. We begin by evaluating model performance on the external mmWave dataset to primarily assess feasibility. We then evaluate on our own collected multi-modal dataset, where we compare 5 GHz and 60 GHz performance and analyze various conditions, including training data size, number of participants, and cross-session generalization.

A. Evaluation on External mmWave Dataset

We first evaluate the feasibility of mmWave-based person identification using the external dataset.

Table I presents the best performance of all the model configurations per model class, chosen w.r.t. the final epoch validation accuracy, in terms of identification test accuracy, with background subtraction enabled. The best-performing configurations achieve strong classification performance, with test accuracies exceeding 95% in several cases, notably for the `OptECAResNet1D-JARIL`, and `LSTM_HumanFi` models. Overall, there is not a lot of variability, and the most accurate model reached a test accuracy of 95.3% $\pm$ 1.5%, showcasing the potential of mmWave CSI for fine-grained biometric identification (using such models).

When not performing background subtraction, for which the results are not shown here due to space constraints, accuracy tends to decrease in general. However, attention-based architectures still retain high performance, with test accuracies remaining above 90% in most cases. This shows that even without background subtraction, our models can still learn to recognize individuals from the raw CSI data.

TABLE I: Person identification accuracy on external mmWave dataset with various models and augmentations. Abbreviations: M = Mixup, GS = Gaussian Smoothing.

Model	Configuration	Data Augmentation	# Parameters	Test Accuracy
CNN-BiLSTM (Temp. Attention)	hidden_dim=128, layers=1	None	207K	0.937 ± 0.051
	hidden_dim=128, layers=2	GS	602K	0.917 ± 0.015
	hidden_dim=64, layers=1	M	74K	0.947 ± 0.042
	hidden_dim=64, layers=2	GS	173K	$\mathbf{0.953 \pm 0.038}$
CNN-BiLSTM (Dual Attention)	hidden_dim=128, layers=1	GS	268K	$\mathbf{0.94 \pm 0.062}$
	hidden_dim=128, layers=2	None	663K	0.917 ± 0.051
	hidden_dim=64, layers=1	GS	86K	0.917 ± 0.042
	hidden_dim=64, layers=2	M	185K	0.92 ± 0.017
CustomECAResNet1D	layers=(1,1,1,1)	M	1.9M	0.94 ± 0.01
CustomResNet1D		None	1.9M	0.937 ± 0.045
OptECAResNet1D-JARIL		M	3.64M	$\mathbf{0.953 \pm 0.023}$
OptResNet1D-JARIL		M	7.03M	0.943 ± 0.047
TemporalConvNet	num_channels=[64, 128], kernel_size=2	M	73K	0.923 ± 0.064
	num_channels=[64, 128, 128], kernel_size=2	GS	139K	0.913 ± 0.021
	num_channels=[64, 128, 256], , kernel_size=2	M + GS	305K	$\mathbf{0.933 \pm 0.046}$
LSTM_HumanFi	hidden_dim=128, layers=1	GS	83K	$\mathbf{0.953 \pm 0.015}$
	hidden_dim=128, layers=2, bidirectional	M + GS	561K	0.937 ± 0.045
	hidden_dim=64, layers=1	GS	25K	0.937 ± 0.015
	hidden_dim=64, layers=2, bidirectional	M + GS	149K	0.93 ± 0.044

B. Results on Our Collected Dataset

Next, we show the results on our own collected dataset, which includes synchronized 5 GHz and 60 GHz CSI measurements from 20 participants. We compare performance between the sub-6 GHz and mmWave modalities, using identical models and training procedures where possible.

We evaluate all models on three CSI configurations: low-rate 5 GHz CSI at a downsampled rate of 10 Hz, high-rate 5 GHz CSI at 200 Hz, and 60 GHz mmWave CSI at a rate of 10 Hz. We report the results with and without background subtraction to assess its impact. All experiments are done using the same participant train-val-test split and fixed five-second sample durations to ensure consistency across models and frequency settings.

1) Sub-6 GHz vs. 60 GHz Comparison Without Background Subtraction: Table II summarizes the best-performing configuration for each model class across three frequency settings, low-rate 5 GHz, high-rate 5 GHz, and low-rate 60 GHz, without background subtraction. The table is horizontally divided into three main sections: the first highlights the top configurations selected based on performance of low-rate 5 GHz, the second section lists the top models for high-rate 5 GHz, and the final section shows those best suited for 60 GHz CSI. For each configuration, we report the test accuracy for all three frequencies to illustrate how well models perform across them, even when originally optimized for another. We remove duplicate models across sections to avoid redundancy, that is the reason why not all sections have 8 models. This structure also applies to Table III.

We observe that increasing the sampling rate to 200 Hz significantly boosts performance for sub-6 GHz CSI across most models, confirming the value of higher sampling rate for capturing motion dynamics. However, the long input sequences (1000 time steps from 200 Hz over 5 seconds) of the high-rate 5 GHz appear to pose challenges for some LSTM-based models, potentially due to difficulties in capturing long-range dependencies or the risk of vanishing gradients over extended temporal contexts.

On 60 GHz data, performance varies a lot across architectures. Although some models (e.g., `TemporalConvNet`) achieve relatively high accuracy, it is still quite lower than for the other frequencies. Other models perform inconsistently and exhibit high variance. Based on the 60 GHz results in Section VI-B2, we primarily attribute this performance gap to the absence of background subtraction, with a more detailed explanation provided in that section.

2) Sub-6 GHz vs. 60 GHz Comparison With Background Subtraction: Table III reports the best person identification accuracy for all model classes across the three frequency settings with background subtraction applied this time.

We can see that, overall, background subtraction leads to improved or more stable accuracy for mmWave CSI (negligible for 5 GHz), for which most models exhibit high classification accuracy, often matching or surpassing their per-

TABLE II: Person identification accuracy comparison across frequencies without background subtraction. Abbreviations: M = Mixup, GS = Gaussian Smoothing, BiDi = Bidirectional, DR = Dropout Rate.

Model	Configuration	#Params	5 GHz @10Hz	5 GHz @200Hz	60 GHz @10Hz
CustomResNet1D	layers=[1,1,1,1], M	1.9M	$\mathbf{0.89 \pm 0.017}$	0.957 ± 0.012	0.293 ± 0.168
CustomECAResNet1D	layers=[1,1,1,1], M	1.9M	$\mathbf{0.897 \pm 0.055}$	0.947 ± 0.006	0.43 ± 0.135
TemporalConvNet	[64,128], kernel_size=2, DR=0.5, M	79K	$\mathbf{0.91 \pm 0.02}$	0.943 ± 0.025	0.883 ± 0.038
LSTM_HumanFi	hidden_dim=128, layers=2, DR=0.2, BiDi, GS	587K	$\mathbf{0.843 \pm 0.023}$	0.62 ± 0.036	0.6 ± 0.062
CNN-BiLSTM (Temp. Attention)	hidden_dim=64, layers=1, GS	79K	$\mathbf{0.83 \pm 0.03}$	0.833 ± 0.067	0.827 ± 0.04
CNN-BiLSTM (Dual Attention)	lstm_units=64, layers=2, GS	188K	$\mathbf{0.81 \pm 0.035}$	0.677 ± 0.042	0.793 ± 0.085
OptECAResNet1D-JARIL	layers=[1,1,1,1], GS	3.64M	$\mathbf{0.857 \pm 0.029}$	0.957 ± 0.012	0.840 ± 0.113
OptResNet1D-JARIL	layers=[1,1,1,1], GS	7.07M	0.83 ± 0.036	0.913 ± 0.032	$\mathbf{0.867 \pm 0.071}$
CustomResNet1D	layers=[1,1,1,1], M + GS	1.9M	$\mathbf{0.893 \pm 0.021}$	0.953 ± 0.029	0.3 ± 0.155
CustomECAResNet1D	layers=[1,1,1,1], M + GS	1.9M	$\mathbf{0.847 \pm 0.021}$	0.92 ± 0.03	0.383 ± 0.071
TemporalConvNet	[64,128,128], kernel_size=3, DR=0.2, M	210K	$\mathbf{0.897 \pm 0.006}$	0.937 ± 0.015	0.857 ± 0.012
LSTM_HumanFi	hidden_dim=128, layers=1, DR=0.5, BiDi	191K	$\mathbf{0.84 \pm 0.03}$	0.733 ± 0.055	0.647 ± 0.051
CNN-BiLSTM (Temp. Attention)	hidden_dim=128, layers=2, M	611K	$\mathbf{0.83 \pm 0.035}$	0.917 ± 0.04	0.77 ± 0.046
CNN-BiLSTM (Dual Attention)	lstm_units=64, layers=1, GS	89K	$\mathbf{0.867 \pm 0.006}$	0.75 ± 0.066	0.813 ± 0.060
OptECAResNet1D-JARIL	layers=[1,1,1,1]	3.64M	$\mathbf{0.723 \pm 0.061}$	0.957 ± 0.032	0.663 ± 0.216
OptResNet1D-JARIL	layers=[1,1,1,1], M + GS	7.07M	$\mathbf{0.76 \pm 0.035}$	0.917 ± 0.025	0.39 ± 0.076
CustomResNet1D	layers=[1,1,1,1]	1.9M	$\mathbf{0.857 \pm 0.045}$	0.953 ± 0.015	0.833 ± 0.110
CustomECAResNet1D	layers=[1,1,1,1], GS	1.9M	$\mathbf{0.867 \pm 0.006}$	0.963 ± 0.012	0.387 ± 0.202
TemporalConvNet	[64,128], kernel_size=3, DR=0.5, M	113K	$\mathbf{0.92 \pm 0.035}$	0.95 ± 0.017	0.887 ± 0.032
LSTM_HumanFi	hidden_dim=64, layers=2, BiDi, DR=0.5, M	166K	$\mathbf{0.84 \pm 0.017}$	0.647 ± 0.045	0.807 ± 0.045
CNN-BiLSTM (Temp. Attention)	hidden_dim=64, layers=2, M + GS	180K	$\mathbf{0.81 \pm 0.05}$	0.903 ± 0.015	0.690 ± 0.137
CNN-BiLSTM (Dual Attention)	lstm_units=128, layers=2	669K	$\mathbf{0.833 \pm 0.031}$	0.677 ± 0.11	0.823 ± 0.045
OptResNet1D-JARIL	layers=[1,1,1,1]	7.07M	$\mathbf{0.803 \pm 0.006}$	0.93 ± 0.036	0.46 ± 0.43

formance at high-rate 5 GHz. Notably, `TemporalConvNet` and `CNN-BiLSTM (Temp. Attention)` both achieve accuracies of 93–96% on 60 GHz data, showing their ability to benefit from spatial detail. In contrast to the no-background-subtraction scenario, models like `CustomResNet1D` and `CustomECAResNet1D` also perform well at 60 GHz, showing that background removal plays an important role here.

The notable improvement in 60 GHz accuracy with background subtraction might be because of the high spatial sensitivity of mmWave signals. Due to their shorter wavelength, mmWave CSI can capture fine-grained (environmental) details, including static clutter that can overwhelm motion-related patterns. Background subtraction may remove this noise, making person-specific features more distinguishable.

In contrast, background subtraction has seemingly little to no benefit for 5 GHz CSI in this case, and can even slightly degrade performance. The reason for this may be that the longer wavelength of the sub-6 GHz signals smooth out fine background details, making static clutter less overwhelming here, and background removal less impactful.

Interestingly, several (LSTM-based) architectures (e.g., `LSTM_HumanFi`, `CNN-BiLSTM (Dual Attention)`) exhibit significantly improved 60 GHz performance compared to their 5 GHz counterparts, especially 5 GHz at 200 Hz, indicating better adaptation to mmWave signal characteristics.

3) Aggregate Frequency Comparison Analysis: To evaluate the overall performance difference between mmWave (60 GHz) and traditional sub-6 GHz (5 GHz) Wi-Fi signals, we also did a comprehensive analysis across all trained model configurations, not only the best ones, using background-subtracted data.

Table IV summarizes the results, showing both average accuracies and counts of configurations where 60 GHz outperforms 5 GHz. Across all models, 60 GHz achieved the highest average accuracy (0.9121), outperforming low-rate 5 GHz in 157 out of 160 configurations, and the high-rate band in 130 configurations.

To assess statistical significance, we used a stricter condition: 60 GHz was considered significantly better when its mean accuracy minus one standard deviation exceeded the 5 GHz accuracy plus its standard deviation. Under this criterion, 60 GHz significantly outperformed 5 GHz @10Hz in 108 configurations, and 5 GHz @200Hz in 105 configurations.

We further break down the results by excluding specific model categories. When removing all LSTM-HumanFi configurations (leaving 96 total configurations), the trend persists: 60 GHz still shows the best overall performance (0.9148 average accuracy) and remains statistically superior in most comparisons. Even when excluding all LSTM-based models (leaving 64 configurations), 60 GHz maintains its lead in

TABLE III: Person identification accuracy comparison across frequencies with background subtraction. Abbreviations: M = Mixup, GS = Gaussian Smoothing, BiDi = Bidirectional, DR = Dropout Rate.

Model	Configuration	#Params	5 GHz @10Hz	5 GHz @200Hz	60 GHz @10Hz
CustomResNet1D	layers=[1,1,1,1], M	1.9M	$\mathbf{0.907 \pm 0.021}$	0.923 ± 0.029	0.85 ± 0.017
CustomECAResNet1D	layers=[1,1,1,1], M	1.9M	$\mathbf{0.897 \pm 0.012}$	0.943 ± 0.006	0.887 ± 0.012
TemporalConvNet	[64,128], kernel_size=2, DR=0.2, M	79K	0.91 ± 0.017	0.933 ± 0.012	$\mathbf{0.963 \pm 0.006}$
LSTM_HumanFi	hidden_dim=64, layers=1, DR=0.2, BiDi	63K	0.887 ± 0.015	0.647 ± 0.029	$\mathbf{0.893 \pm 0.012}$
CNN-BiLSTM (Temp. Attention)	hidden_dim=64, layers=1, M + GS	79K	0.817 ± 0.021	0.87 ± 0.03	$\mathbf{0.947 \pm 0.012}$
CNN-BiLSTM (Dual Attention)	lstm_units=64, layers=1, GS	89K	0.843 ± 0.015	0.647 ± 0.065	$\mathbf{0.9 \pm 0.017}$
OptECAResNet1D-JARIL	layers=[1,1,1,1], GS	3.64M	0.86 ± 0.05	0.933 ± 0.042	$\mathbf{0.903 \pm 0.031}$
OptResNet1D-JARIL	layers=[1,1,1,1], GS	7.06M	0.85 ± 0.035	0.853 ± 0.075	$\mathbf{0.907 \pm 0.049}$
CustomResNet1D	layers=[1,1,1,1], GS	1.9M	0.853 ± 0.040	0.893 ± 0.074	$\mathbf{0.947 \pm 0.012}$
CustomECAResNet1D	layers=[1,1,1,1], GS	1.9M	0.847 ± 0.012	0.943 ± 0.006	$\mathbf{0.937 \pm 0.006}$
TemporalConvNet	[64,128,256], kernel_size=2, DR=0.2, M + GS	312K	0.86 ± 0.017	0.953 ± 0.015	$\mathbf{0.923 \pm 0.032}$
LSTM_HumanFi	hidden_dim=64, layers=1, DR=0.5, BiDi	63K	0.863 ± 0.006	0.657 ± 0.068	$\mathbf{0.913 \pm 0.015}$
CNN-BiLSTM (Dual Attention)	lstm_units=128, layers=1	273K	0.847 ± 0.021	0.753 ± 0.040	$\mathbf{0.91 \pm 0.03}$
OptResNet1D-JARIL	layers=[1,1,1,1]	7.06M	0.747 ± 0.041	0.963 ± 0.012	$\mathbf{0.853 \pm 0.015}$
CustomResNet1D	layers=[1,1,1,1], M + GS	1.9M	0.9 ± 0.0	0.937 ± 0.015	$\mathbf{0.92 \pm 0.046}$
CustomECAResNet1D	layers=[1,1,1,1], M + GS	1.9M	0.883 ± 0.050	0.93 ± 0.026	$\mathbf{0.907 \pm 0.006}$
TemporalConvNet	[64,128,128], kernel_size=2, DR=0.5, M	147K	0.93 ± 0.0	0.943 ± 0.025	$\mathbf{0.947 \pm 0.012}$
LSTM_HumanFi	hidden_dim=128, layers=1, BiDi, DR=0.2, GS	200K	0.847 ± 0.021	0.623 ± 0.032	$\mathbf{0.953 \pm 0.012}$
CNN-BiLSTM (Temp. Attention)	hidden_dim=128, layers=2, M + GS	611K	0.837 ± 0.035	0.87 ± 0.044	$\mathbf{0.93 \pm 0.03}$
CNN-BiLSTM (Dual Attention)	lstm_units=128, layers=1, M + GS	273K	0.817 ± 0.012	0.727 ± 0.057	$\mathbf{0.933 \pm 0.006}$
OptECAResNet1D-JARIL	layers=[1,1,1,1]	3.64M	0.767 ± 0.025	0.953 ± 0.015	$\mathbf{0.897 \pm 0.038}$

TABLE IV: Comparison of 60 GHz vs. 5 GHz performance across all model configurations. Significant improvement ($\gg$) is defined as: $\mu_{60GHz} - \sigma_{60GHz} > \mu_{5GHz} + \sigma_{5GHz}$.

Metric	All Models (160)	Excl. LSTM-HumanFi (96)	Excl. All LSTM-based (64)
Avg. Accuracy (60 GHz)	**0.9121**	**0.9148**	**0.9153**
Avg. Accuracy (5 GHz @10Hz)	0.8432	0.8444	0.8584
Avg. Accuracy (5 GHz @200Hz)	0.7526	0.8536	0.9010
60 GHz $>$ 5 GHz @10Hz (# models)	**157**	**93**	**61**
60 GHz $>$ 5 GHz @200Hz (# models)	**130**	**66**	**34**
60 GHz $\gg$ 5 GHz @10Hz (1 std)	**108**	**69**	**41**
60 GHz $\gg$ 5 GHz @200Hz (1 std)	**105**	41	18

average accuracy (0.9153), although 5 GHz at 200 Hz becomes more competitive (and dominant), particularly for CNN-based models.

These findings support the conclusion that for a comparable sampling rate, mmWave CSI outperforms traditional Wi-Fi sensing for person identification, especially when considering statistical robustness. However, high-rate sub-6 GHz data can beat mmWave performance under a decent number of model configurations, indicating that both resolution and temporal density play key roles in system accuracy.

It is important to note that all configurations compared here and in Table III use background-subtracted data. Although this preprocessing step improves performance for 60 GHz CSI, it has little effect, or even slightly negative impact, on 5 GHz accuracy as we noticed. We can also implicitly look at 60 GHz with background subtraction against non-subtracted 5 GHz data using the two tables. Even in this comparison, 60 GHz still consistently outperforms low-rate 5 GHz. However, against high-rate 5 GHz, mmWave performance becomes more competitive but not really dominant, especially when considering non-LSTM-based architectures.

In summary, background subtraction substantially boosts the performance of mmWave CSI across models, allowing it to clearly outperform low-rate 5 GHz and approach the accuracy of high-rate 5 GHz. However, high-rate 5 GHz, particularly without background subtraction, can still surpass mmWave performance in certain model classes, especially CNN-based architectures.

4) Learning Curve Analysis: To assess how frequency performance scales with training data size, we conduct a

learning curve experiment. We select a representative model that demonstrates consistently strong performance across all frequency settings: the `TemporalConvNet` architecture, configured with three temporal convolutional blocks of widths (filters) $[64, 128, 128]$, a kernel size of 2, dropout rate of 0.5, and trained with Mixup data augmentation.

This model is trained on progressively smaller portions of the training data (i.e., 70%, 60%, ..., down to 10% of the full dataset), while performance is evaluated on a fixed test set (15% of the total dataset). The experiment is repeated for sample durations of 2, 3, and 5 seconds to also evaluate how temporal context influences identification accuracy. For consistency, the same model configuration is also used in Sections VI-B5 and VI-B6.

Fig. 4 shows the learning curves for the three different sample duration values. The 60 GHz data consistently outperforms both 5 GHz variants, showing better robustness to reduced training data. The performance gap increases significantly at smaller training fractions (below 30%). 60GHz maintains its advantage whether we use short or long data sample duration values, meaning that this benefit does not depend on sample duration.

5) Effect of Number of Participants: We also evaluate how well the frequencies generalize as the number of target classes (i.e., participants) decreases in the total dataset. We use subsets of 5, 10, 15, and 20 participants, and evaluate performance at each level. Note that when we use a subset that contains less than 20 participants, they are picked randomly from the total pool of participants. Since we run this for multiple seeded runs, as we do everywhere, the random subset is also different for every seed. This is again repeated for multiple sample durations.

Fig. 5 shows how person identification accuracy scales with the number of participants. For 2-second and 3-second sample durations, 60 GHz shows comparable or superior performance compared to 5 GHz, especially at higher participant counts (10–20 people). For the 5-second sample case, 60 GHz performs worse overall when compared to 5 GHz, except when the participant count reaches 20. Regarding 5GHz variants, 10Hz only outperforms 200Hz at 5 participants, with 200Hz demonstrating better accuracy for larger groups.

This suggests that 60 GHz CSI becomes more advantageous compared to 5GHz as the number of individuals grows, likely due to its ability to capture subtle spatial patterns that distinguish many similar gaits. However, for smaller group sizes, simpler signals with lower resolution (i.e., 5 GHz) may suffice.

6) Cross-Session Evaluation: To test temporal robustness, we try a cross-session evaluation where training and testing data come from different collection days. In our dataset, there are two participants for whom we collected data on two different days. We want to see whether our model, trained on a certain day for those participants, can recognize their sequence if it were from another day.

As shown in Fig. 6, all models fail to generalize across sessions: classification accuracy drops to near 0% for both participants across all frequency bands. The 60 GHz model is seemingly kind of biased, consistently predicting a single class that is present in the same day session as the participants' test data, which is maybe somewhat understandable. Meanwhile, predictions from the 5 GHz models (at both 10 Hz and 200 Hz) appear erratic, showing no clear bias toward a specific incorrect class. All of this, in general, further emphasizes the challenge of generalizing over temporal (and possibly environmental) shifts in CSI data.

VII. Limitations

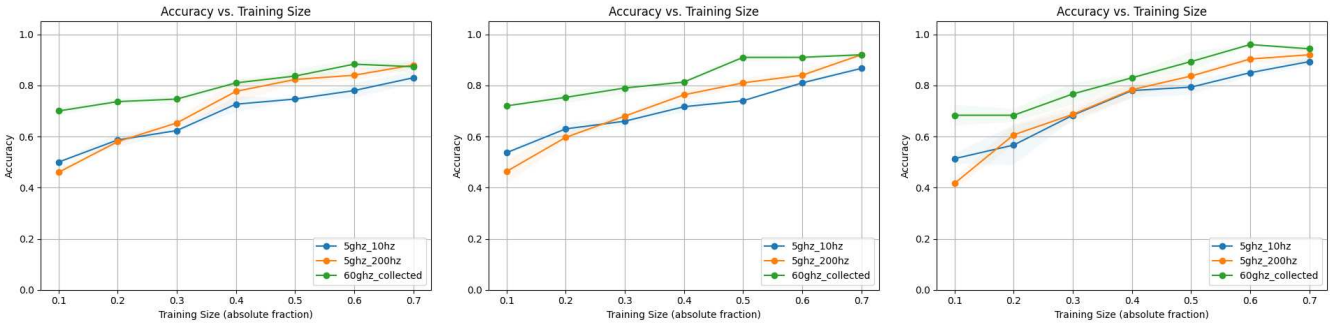
In this section, we discuss some problems or limitations encountered during this study. First, while mmWave CSI achieved relatively high person identification accuracy within our experimental setup, the specific physical placement of the mmWave transceivers, arranged in an 'X'-shaped configuration, may not necessarily be optimal. Specifically, at the start and end of the walking path, the person does not block the signal path between the transceivers. As a result, the corresponding portions of the CSI sequence carry little to no person-specific information. This issue was noted in a related localization study using the same setup, which used a 5×4 grid of positions within the sensing area.

Second, for the 5 GHz dataset collected at 200 Hz (i.e., 200 packets per second), we were only able to include data from 18 participants, rather than the full set of 20. This was due to two individuals having unexpectedly low sampling rates ($\sim$30 Hz) during data collection, making it impossible to upsample their samples to the target rate. Although prior work suggests that reducing the subject count from 20 to 18 is unlikely to drastically impact classification performance, we acknowledge this lack of total fairness.

Finally, a minor oversight lies in the way background subtraction was used. Namely, the background sequence was collected only on the third day of the CSI data recordings, as mentioned previously. This assumes relative stability in the static channel conditions, which may hold approximately for sub-6 GHz Wi-Fi due to its longer channel coherence times in non-changing indoor environments. However, for mmWave, the channel is highly sensitive to even minor special environmental variations, such as air movement, humidity, or thermal changes, because of its much shorter wavelength [59]. As a result, reusing background samples from a different session can introduce inconsistencies, especially in the mmWave data. Ideally, background calibration should be performed for each recording session, or even more frequently in practical deployments, to better match the (instantaneous) changing channel conditions. [60]

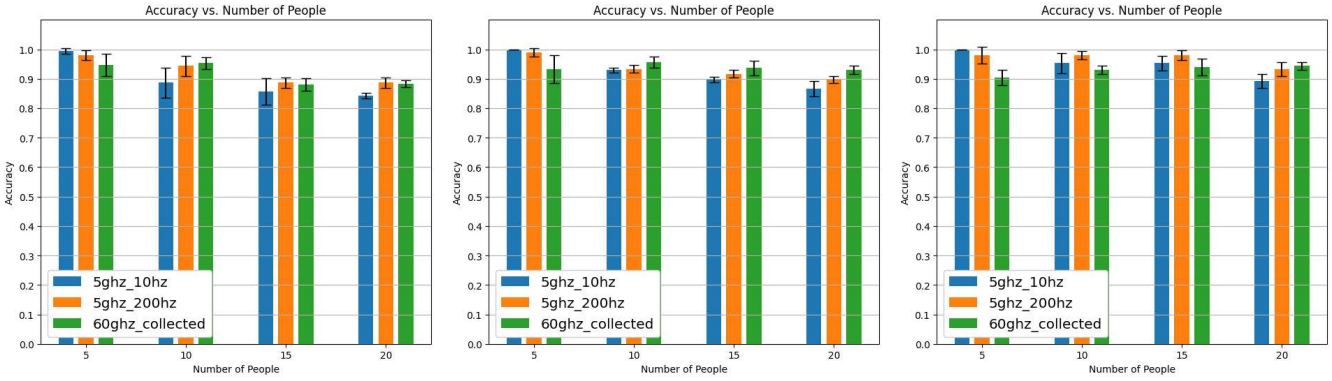
VIII. Conclusion and Future Work

In this study, we explored person identification using Wi-Fi-based sensing at both 5 GHz and 60 GHz frequencies through a newly collected dataset of synchronized measurements from 20 participants. Our multi-frequency dataset, captured using COTS hardware in a controlled environment, enabled direct comparison between frequency bands under



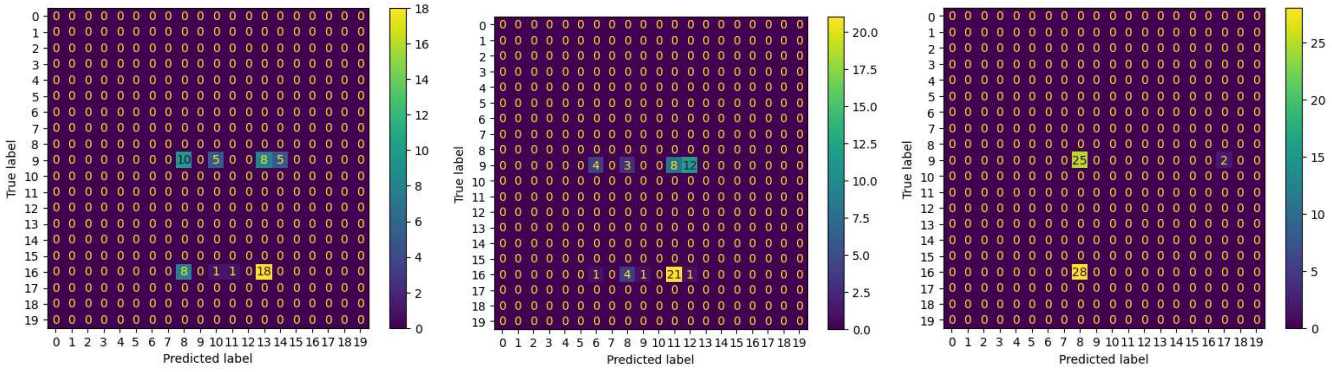
(a) Sample duration: 2 seconds (b) Sample duration: 3 seconds (c) Sample duration: 5 seconds

Fig. 4: Learning curves for the `TemporalConvNet` model at three different sample durations. Each plot shows test accuracy as a function of the amount of training data used (from 10% to 70%), for 5 GHz @10 Hz (blue), 5 GHz @200 Hz (orange), and 60 GHz CSI signals (green).



(a) Sample duration: 2 seconds (b) Sample duration: 3 seconds (c) Sample duration: 5 seconds

Fig. 5: Effect of participant count on frequency accuracy for the `TemporalConvNet` model at three different sample durations. Each plot shows test accuracy as a function of the number of people used in the dataset, for 5 GHz @10 Hz (blue), 5 GHz @200 Hz (orange), and 60 GHz CSI signals (green).



(a) 5 GHz @10 Hz (b) 5 GHz @200 Hz (c) 60 GHz @10 Hz

Fig. 6: Confusion matrices for the `TemporalConvNet` model for cross-session evaluation on two participants. All frequencies show near-zero identification accuracy when models are trained and tested on data from different days.

identical conditions. The experiments showed that 60 GHz signals achieve superior accuracy (91.2% average) when combined with background subtraction, demonstrating particular advantages for larger participant groups and limited training data scenarios. However, high-rate 5 GHz sampling proved equally competitive in many cases, suggesting that temporal resolution can compensate to a certain extent for spatial resolution limitations. The performance hierarchy reversed without background subtraction, highlighting mmWave's sensitivity to environmental clutter. We also identified critical limitations regarding cross-session generalization.

Future work could investigate several directions. First, we plan to explore how mmWave device placement and quantity affect identification performance, which was referenced in Section VII. Second, we will investigate into advanced domain adaptation techniques to improve cross-session performance, examining whether mmWave's finer spatial resolution can somehow provide an advantage in maintaining identity signatures across different sessions. Finally, we intend to investigate mmWave sensing with higher sampling rates, as our used sampling rate was constrained by the device-specific firmware. Overcoming this constraint could certainly unlock further performance gains.

IX. Acknowledgments

This work was carried out under the supervision of Prof. Jeroen Famaey, whose guidance is gratefully acknowledged. The authors would especially like to express their appreciation to mentors Nabeel Nisar Bhat and Jakob Struye for their practical support and valuable input throughout the project. Finally, sincere thanks are extended to all participants involved in the data collection process for their time and contribution.

References

[1] Kiran B. Raja, Egidijus Auksorius, R. Raghavendra, A. Claude Boccara, and Christoph Busch. Robust verification with subsurface fingerprint recognition using full field optical coherence tomography. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[2] Hyun-Ae Park and Kang Ryoung Park. Iris recognition based on score level fusion by using svm. *Pattern Recognition Letters*, 28(15):2019–2028, 2007.

[3] Claudia Nickel, Christoph Busch, Sathyanarayanan Rangarajan, and Manuel Möbius. Using hidden markov models for accelerometer-based biometric gait recognition. In *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, pages 58–63, 2011.

[4] Thanh Trung Ngo, Yasushi Makihara, Hajime Nagahara, Yasuhiro Mukaigawa, and Yasushi Yagi. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, 47(1):228–237, 2014.

[5] Jing Li, Tao Qiu, Chang Wen, Kai Xie, and Fang-Qing Wen. Robust face recognition using the deep c2d-cnn model based on decision-level fusion. *Sensors*, 18(7), 2018.

[6] Yang Feng, Yuncheng Li, and Jiebo Luo. Learning effective gait features using lstm. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 325–330, 2016.

[7] Reza Shahbazian and Irina Trubitsyna. Human sensing by using radio frequency signals: A survey on occupancy and activity detection. *IEEE Access*, 11:40878–40904, 2023.

[8] Yongsen Ma, Gang Zhou, and Shuangquan Wang. Wifi sensing with channel state information: A survey. *ACM Comput. Surv.*, 52(3), June 2019.

[9] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. Wiwho: Wifi-based person identification in smart spaces. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 1–12, 2016.

[10] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: gathering 802.11n traces with channel state information. *SIGCOMM Comput. Commun. Rev.*, 41(1):53, January 2011.

[11] Glenn Forbes. Csikit: Python csi processing and visualisation tools for commercial off-the-shelf hardware., 2021.

[12] Glenn Forbes, Stewart Massie, and Susan Craw. Wifi-based human activity recognition using raspberry pi. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 722–730, 2020.

[13] Francesco Gringoli, Marco Cominelli, Alejandro Blanco, and Joerg 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*, WiNTECH '21, page 46–53, New York, NY, USA, 2021. Association for Computing Machinery.

[14] Christopher Vaughan, Brian Davis, and Jeremy OCononor. *Dynamics of human gait*. South Africa: Kiboho Publishers, 2nd edition, 1999.

[15] Mary Pat Murray, A. B. Drought, and Ross C. Kory. Walking patterns of normal men. *The Journal of bone and joint surgery. American volume*, 1964.

[16] Mary Pat Murray. Gait as a total pattern of movement. *American Journal of PhysicalMedicine*, 46(1):290–333, 1967.

[17] Jin Zhang, Bo Wei, Wen Hu, and Salil S. Kanhere. Wifi-id: Human identification using wifi signal. In *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 75–82, 2016.

[18] Wei Wang, Alex X. Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, page 363–373, New York, NY, USA, 2016. Association for Computing Machinery.

[19] Akarsh Pokkunuru, Kalvik Jakkala, Arupjyoti Bhuyan, Pu Wang, and Zhi Sun. Neuralwave: Gait-based user identification through commodity wifi and deep learning. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, pages 758–765, 2018.

[20] Xingxia Ming, Hongwei Feng, Qirong Bu, Jing Zhang, Gang Yang, and Tuo Zhang. Humanfi: Wifi-based human identification using recurrent neural network. In *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 640–647, 2019.

[21] Jin Zhang, Bo Wei, Fuxiang Wu, Limeng Dong, Wen Hu, Salil S. Kanhere, Chengwen Luo, Shui Yu, and Jun Cheng. Gate-id: Wifi-based human identification irrespective of walking directions in smart home. *IEEE Internet of Things Journal*, 8(9):7610–7624, 2021.

[22] Tianbo Gu, Zheng Fang, Zhicheng Yang, Pengfei Hu, and Prasant Mohapatra. mmsense: Multi-person detection and identification via mmwave sensing. In *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, mmNets '19, page 45–50, New York, NY, USA, 2019. Association for Computing Machinery.

[23] Nabeel Nisar Bhat, Rafael Berkvens, and Jeroen Famaey. Gesture recognition with mmwave wi-fi access points: Lessons learned. In *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 127–136, 2023.

[24] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 3d tracking via body radio reflections. In *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, NSDI'14, page 317–329, USA, 2014. USENIX Association.

[25] Fadel Adib and Dina Katabi. See through walls with wifi! *SIGCOMM Comput. Commun. Rev.*, 43(4):75–86, August 2013.

[26] Yunze Zeng, Parth H. Pathak, Chao Xu, and Prasant Mohapatra. Your ap knows how you move: fine-grained device motion recognition through wifi. In *Proceedings of the 1st ACM Workshop on Hot Topics in Wireless*, HotWireless '14, page 49–54, New York, NY, USA, 2014. Association for Computing Machinery.

[27] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, Mobi-

Com '14, page 617–628, New York, NY, USA, 2014. Association for Computing Machinery.

[28] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. Analyzing shopper's behavior through wifi signals. In *Proceedings of the 2nd Workshop on Workshop on Physical Analytics*, WPA '15, page 13–18, New York, NY, USA, 2015. Association for Computing Machinery.

[29] Yuxi Wang, Kaishun Wu, and Lionel M. Ni. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594, 2017.

[30] Jinyang Huang, Bin Liu, Chao Chen, Hongxin Jin, Zhiqiang Liu, Chi Zhang, and Nenghai Yu. Towards anti-interference human activity recognition based on wifi subcarrier correlation selection. *IEEE Transactions on Vehicular Technology*, 69(6):6739–6754, 2020.

[31] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, MobiCom '13, page 27–38, New York, NY, USA, 2013. Association for Computing Machinery.

[32] Jianfei Yang, Han Zou, Yuxun Zhou, and Lihua Xie. Learning gestures from wifi: A siamese recurrent convolutional architecture. *IEEE Internet of Things Journal*, 6(6):10763–10772, 2019.

[33] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. Wi-sleep: Contactless sleep monitoring via wifi signals. In *2014 IEEE Real-Time Systems Symposium*, pages 346–355, 2014.

[34] Jin Zhang, Weitao Xu, Wen Hu, and Salil S. Kanhere. Wicare: Towards in-situ breath monitoring. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MobiQuitous 2017, page 126–135, New York, NY, USA, 2017. Association for Computing Machinery.

[35] Kamran Ali, Alex X. Liu, Wei Wang, and Muhammad Shahzad. Keystroke recognition using wifi signals. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 90–102, New York, NY, USA, 2015. Association for Computing Machinery.

[36] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. Csi-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology*, 66(1):763–776, 2017.

[37] Xuyu Wang, Xiangyu Wang, and Shiwen Mao. Cifi: Deep convolutional neural networks for indoor localization with 5 ghz wi-fi. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6, 2017.

[38] Zhongfei Ni and Binke Huang. Gait-based person identification and intruder detection using mm-wave sensing in multi-person scenario. *IEEE Sensors Journal*, 22(10):9713–9723, 2022.

[39] Dongjiang Cao, Ruofeng Liu, Hao Li, Shuai Wang, Wenchao Jiang, and Chris Xiaoxuan Lu. Cross vision-rf gait re-identification with low-cost rgb-d cameras and mmwave radars. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(3), September 2022.

[40] Jian Guo, Jingpeng Wei, Yashan Xiang, and Chong Han. Millimeter-wave radar-based identity recognition algorithm built on multimodal fusion. *Sensors*, 24(13), 2024.

[41] Alejandro Blanco, Pablo Jiménez Mateo, Francesco Gringoli, and Joerg Widmer. Augmenting mmwave localization accuracy through sub-6 ghz on off-the-shelf devices. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, MobiSys '22, page 477–490, New York, NY, USA, 2022. Association for Computing Machinery.

[42] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, New York, NY, USA, 2005. The standard textbook for MIMO and wireless channel modeling.

[43] *Signal Processing Fundamentals for JCAS*, chapter 2, pages 31–58. John Wiley Sons, Ltd, 2022.

[44] Jaeseong Son and Jaesung Park. Channel state information (csi) amplitude coloring scheme for enhancing accuracy of an indoor occupancy detection system using wi-fi sensing. *Applied Sciences*, 14(17), 2024.

[45] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[46] Ilaria Cacciari and Anedio Ranfagni. Hands-on fundamentals of 1d convolutional neural networks—a tutorial for beginner users. *Applied Sciences*, 14(18), 2024.

[47] Gianni Brauwers and Flavius Frasincar. A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, PP:1–1, 11 2021.

[48] Jörg Schäfer, Baldev Raj Barrsiwal, Muyassar Kokhkharova, Hannan Adil, and Jens Liebehenschel. Human activity recognition using csi information with nexmon. *Applied Sciences*, 11(19), 2021.

[49] Matthias Schulz, Daniel Wegemer, and Matthias Hollick. Nexmon: The c-based firmware patching framework, 2017.

[50] Francesco Gringoli, Matthias Schulz, Jakob Link, and Matthias 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*, WiNTECH '19, page 21–28, 2019.

[51] IMDEA Networks. Mikrotik-researcher-tools: Openwrt firmware for mmwave sensing experiments. https://github.com/IMDEANetworksWNG/Mikrotik-researcher-tools, 2021.

[52] Jakob Struye. 60gx3_scripts: Data collection scripts for 60 ghz wifi sensing experiments. https://github.com/JakobStruye/60gx3_scripts, 2024.

[53] David A. Forsyth and Jean Ponce. *Computer Vision - A Modern Approach, Second Edition.* Pitman, 2012. pages 407–409.

[54] Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *CoRR*, abs/1710.09412, 2017.

[55] Zhenyu Zhu, Qing Yang, Xin Liu, and Dexin Gao. Attention-based cnn-bilstm for soh and rul estimation of lithium-ion batteries. *Journal of Algorithms Computational Technology*, 16:174830262211305, 10 2022.

[56] Fei Wang, Jianwei Feng, Yinliang Zhao, Xiaobin Zhang, Shiyuan Zhang, and Jinsong Han. Joint activity recognition and indoor localization with wifi fingerprints. *IEEE Access*, 7:80058–80068, 2019.

[57] Pengfei Zhu Peihua Li Wangmeng Zuo Qilong Wang, Banggu Wu and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[58] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv:1803.01271*, 2018.

[59] Theodore S. Rappaport, Robert W. Heath, Robert C. Daniels, and James N. Murdock. *Millimeter Wave Wireless Communications*. Pearson Education, 2014. pages 1–32.

[60] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge, UK, 2005. pages 21–56.

[61] Belal Korany, Hong Cai, and Yasamin Mostofi. Multiple people identification through walls using off-the-shelf wifi. *IEEE Internet of Things Journal*, 8(8):6963–6974, 2021.

[62] Belal Korany, Chitra R. Karanam, Hong Cai, and Yasamin Mostofi. Xmodal-id: Using wifi for through-wall person identification from candidate video footage. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[63] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.