One-Sided CSI-Based Sensing in Adversarial Through-Wall Settings

VYTAUTAS BAKANAS, University of Twente, The Netherlands

Wi-Fi signals can reveal patterns of human activity by analyzing how movement disturbs signal propagation. While prior work on Channel State Information (CSI) has shown promise for indoor activity recognition, it typically assumes sensor placement within the monitored environment or in adjacent rooms. This study investigates a more constrained and privacy-sensitive scenario: can activity inside a home be inferred using low-cost wireless devices placed entirely outside, with no interior access? Using a custom active sensing setup, we collected over 600 minutes of CSI data across two residential apartments, capturing room-level presence under varying wall materials, device placements, and participant behavior. We evaluated multiple learning strategies—including supervised, unsupervised, and semi-supervised models—to assess the feasibility of inferring interior activity from exterior signals. Results show that room-level localization is achievable even through thick residential walls, though performance varies substantially with environmental structure and sensor configuration. These findings demonstrate the viability of a previously unexplored form of passive activity inference and raise important questions about wireless privacy in domestic settings.

Additional Key Words and Phrases: Wi-Fi sensing, channel state information, human activity recognition, through-wall sensing, ESP32, privacy, one-sided, localization, semi-supervised clustering, deep learning

1 INTRODUCTION

Human Activity Recognition (HAR) using Wi-Fi signals has gained significant attention due to its unobtrusive sensing capabilities. Channel State Information, which captures amplitude and phase variations across Wi-Fi subcarriers, has proven particularly effective for detecting fine-grained motion [16]. CSI enables recognition of subtle activities such as breathing [18] or hand gestures [22], and offers privacy advantages over camera-based systems by relying on signal reflections instead of visual data [2].

Wi-Fi CSI sensing has increasingly been proposed as a privacy-preserving alternative to camera-based and wearable activity recognition systems, especially in health and eldercare contexts. Unlike cameras, CSI does not capture visual or personally identifiable data, and unlike wearables, it requires no physical contact or active compliance—making it well-suited for continuous monitoring of individuals who may forget, reject, or be unable to use body-worn devices. This is particularly relevant for older adults and individuals with neurodegenerative conditions, where reliable in-home sensing is critical but intrusive solutions are often unacceptable [5]. Recent studies demonstrate that CSI-based systems can detect subtle human motions—such as tremors or hand gestures—with high temporal precision, enabling non-contact health monitoring in everyday residential settings [6]. These characteristics position CSI sensing as a promising tool for unobtrusive, dignity-preserving monitoring of physical activity in vulnerable populations.

TScIT 43, July 4, 2025, Enschede, The Netherlands

Most existing CSI-based HAR systems rely on sensors placed inside the monitored space, often assuming line-of-sight (LOS) between transmitters and receivers. However, Wi-Fi signals are capable of penetrating walls, raising the possibility of activity inference from outside a building. While earlier studies have investigated through-wall sensing within indoor environments [23], little is known about sensing in fully external, adversarial conditions—where neither the transmitter nor the receiver has access to the interior.

This work explores that scenario directly: can low-cost ESP32 devices placed entirely outside a building detect in which room a person is located, based solely on Wi-Fi Channel State Information? We use an active CSI sensing setup, where an ESP-NOW-enabled transmitter broadcasts packets to a passive receiver placed on the building's exterior. CSI spectrograms derived from the received signals are used to classify spatially localized activity — specifically, presence in the kitchen, hallway, or bathroom — using both supervised deep learning and unsupervised or semi-supervised clustering techniques.

Research question (RQ): To what extent can Wi-Fi CSI collected from outside a building be used to infer room-level occupancy indoors?

To explore this overarching question, we examine the following sub-questions:

- **SRQ1:** How does transmitter orientation (0°, ±45°) and shielding (open vs. boxed) affect classification and clustering accuracy and F1 scores for room-level presence detection?
- **SRQ2:** What classification accuracy and F1 scores can be achieved for room-level human activity recognition using supervised, unsupervised, and semi-supervised models trained on CSI spectrograms?
- **SRQ3:** How does sensing performance (in terms of accuracy and variability) generalize across two different apartment layouts and multiple participants?

This set of questions is relevant both technically and socially, as it explores whether activity inside a private home can be inferred using inexpensive external devices. By challenging the traditional notion that walls guarantee privacy, this research opens up important discussions about the boundaries of domestic space, personal security, and the ethical implications of ambient wireless surveillance.

The remainder of this thesis is structured as follows. Section 2 reviews prior work on human activity recognition using Wi-Fi CSI and related sensing modalities. Section 3 describes the experimental setup and data collection process across two apartment environments. Section 4 details the data preprocessing steps and modeling strategies used for activity classification. Section 5 presents the experimental results, structured by environment, model type, and configuration sensitivity. Section 6 discusses the findings in depth, including the strengths and limitations of the study, and Section 7 concludes with a summary and directions for future research.

2 RELATED WORK

CSI-based Wi-Fi sensing has been widely explored for HAR tasks. Early systems such as E-eyes [27] and CARM [26] demonstrated that variations in wireless propagation, caused by human movement, could be captured using Intel 5300 NICs with custom drivers [12]. These works showed successful activity classification in indoor and LOS scenarios. Subsequent models incorporated deep learning techniques such as convolutional and recurrent neural networks to improve performance [25, 29].

Beyond line-of-sight (LOS) environments, prior research has demonstrated that CSI can detect activity in non-line-of-sight or through-wall scenarios [23]. These studies commonly involve sensor placements inside buildings or in adjacent rooms separated by thin interior walls, allowing relatively direct or low-attenuation signal paths. While these configurations show the feasibility of through-wall sensing, they typically assume some degree of proximity and structural simplicity.

The evolution of CSI-capable hardware has also enabled broader deployment. Intel 5300 NICs offered high-quality CSI but were limited by power and platform constraints. Atheros and Broadcom chipsets later enabled embedded platforms such as Raspberry Pi to be used for CSI collection [11, 28]. Recent developments have made it possible to collect CSI data on ESP32 microcontrollers, which are inexpensive, energy-efficient, and easy to deploy [14, 23].

This study makes four contributions to the field of CSI-based human activity recognition. First, it investigates a rarely explored sensing configuration in which both the transmitter and receiver are positioned on the same side of the wall—such as in a shared hallway or entirely outside the apartment—with no access to the monitored interior. To our knowledge, fully external, one-sided sensing setups have only been reported once [13]. Second, the study collects 648 minutes of labeled CSI data across two distinct apartment environments, enabling a more comprehensive evaluation of model performance. Third, we apply a range of learning strategies, including supervised classification with a CNN [15], unsupervised clustering using K-Means [17] and Gaussian Mixture Models [7], and semi-supervised clustering with MPCK-Means-M-F [4] and COP-KMeans [24] to compare performance across labeled and unlabeled conditions. Finally, we assess how sensing performance varies by apartment layout, wall materials, transmitter orientation, shielding, and participant identity—offering insight into the environmental and human factors that affect CSI-based inference.

3 DATA ACQUISITION

To evaluate the feasibility of one-sided, through-wall human activity recognition using CSI, we conducted a series of structured experiments across two different residential environments. This section describes the hardware setup, experimental design, and data collection protocols used in the study. We outline the roles of the ESP32-C6 [9] devices, the spatial configurations in each apartment, and the procedures used to label room-level activity. Special attention is given to how environmental variation, participant behavior, and potential sensing limitations were managed to ensure consistent and interpretable data.

3.1 Experimental Setup and Hardware

The sensing system consisted of two ESP32-C6 [9] boards configured for active sensing. One acted as a *transmitter*, sending ESP-NOW packets at 20 Hz, while the other functioned as a *passive receiver* connected to a Raspberry Pi 4B via USB. CSI data was extracted using Espressif's open-source *ESP-CSI* toolchain [1], and the Pi recorded timestamped packets for later processing. Both ESP32's were equipped with *ALFA Network APA-M25 directional panel antennas* [3] to enhance signal directivity.

In Apartment A , the transmitter was placed 1 meter in front of the entrance door, and the receiver was positioned 2 metes to the side, mounted on the exterior wall of the apartment, within the shared hallway of the building. The full layout, including sensor placements, is shown in Figure 1a. This parallel arrangement ensured that signal propagation occurred through a solid reinforced concrete wall. To suppress side-path signal leakage and increase sensitivity to interior activity, the receiver was partially shielded with aluminum foil. Example transmitter setup combinations are illustrated in Figure 2.

Six configurations were tested by combining:

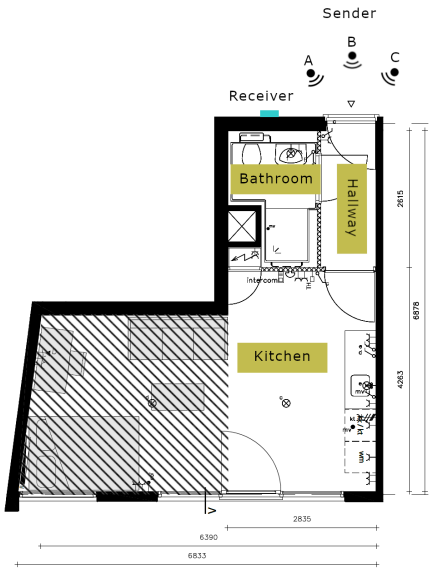
- **Transmitter orientation:** 0° (in front of the door), 45° left, and 45° right
- **Shielding condition:** open-air vs. transmitter enclosed in a cardboard box lined with aluminum foil, with the opening facing the door

A sample of these experimental configurations is shown in Figure 2.

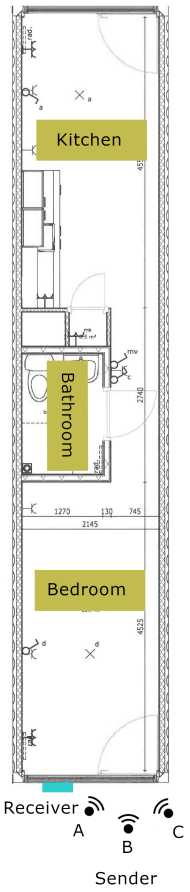
Each configuration in Apartment A was repeated across three participants, with each participant completing two trials per configuration. In each trial, four activity conditions—movement in the kitchen, hallway, and bathroom, and an empty apartment baseline—were recorded. The order of these activities was randomized using the Random.org List Randomizer service [20] to prevent order effects. Each condition lasted for 3 minutes, resulting in a 12-minute CSI recording per trial. With 3 angles × 2 shielding conditions × 2 trials × 3 participants, a total of 36 recording sessions were collected in Apartment A, amounting to 432 minutes of CSI data sampled at 20Hz.

During preliminary calibration, a section of the kitchen in Apartment A was found to be consistently unreachable by any of the tested transmitter-receiver placements. This region, marked in Figure 1a with a hatch pattern, produced CSI signatures indistinguishable from the empty apartment baseline. To avoid contaminating the kitchen class label, participants were explicitly instructed not to enter this dead zone during recording. Consequently, all kitchen activity was constrained to the signal-reachable portion of the room.

To evaluate the reproducibility of the sensing approach in a different spatial context, an additional round of data collection was conducted in a second apartment with a distinct layout and wall materials. Unlike Apartment A, which was a conventional brick-and-concrete residential unit, Apartment B was a prefabricated shipping container apartment. This structure features thinner walls, metal framing, and more prominent glass elements, such as large windows and a partially glazed entrance. Due to spatial constraints, the transmitter in Apartment B was mounted only 1.5 m away from the door, compared to 2 m in Apartment A. The same six transmitter

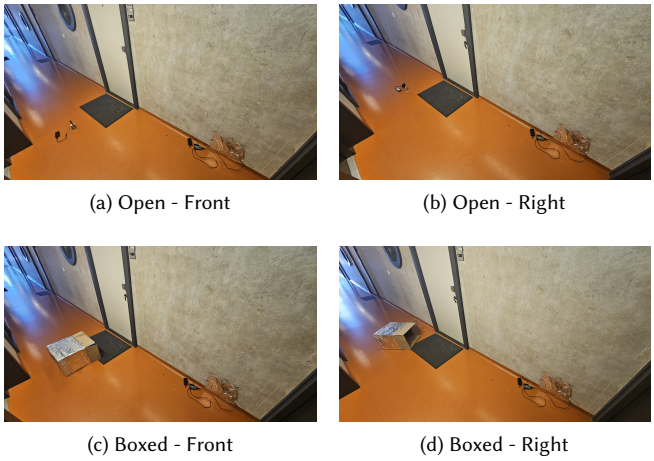


(a) Apartment A: kitchen, hallway, bathroom. Transmitter placements A–C = left/front/right.



(b) Apartment B: kitchen, bedroom, bathroom. Transmitter placements A–C = right/front/left.

Fig. 1. Layouts and sensor placements in Apartment A and Apartment B.



(a) Open - Front (b) Open - Right

(c) Boxed - Front (d) Boxed - Right

Fig. 2. Example transmitter configurations in Apartment A. Rows show shielding condition (open, boxed); columns show orientation (front, right).

configurations were replicated in Apartment B, using identical placement logic, preprocessing steps, and experimental procedures. While Apartment A included labeled regions for kitchen, hallway, and bathroom activities, the layout of Apartment B featured a kitchen, bedroom, and bathroom. Thus, although the kitchen and bathroom classes were preserved across environments, the hallway class in Apartment A was replaced by a bedroom class in Apartment B. All analyses—including supervised classification, unsupervised clustering, and semi-supervised clustering—were repeated using data from the new environment, with models retrained from scratch but the pipeline held constant.

To minimize ambient interference from other residents, all recordings were conducted on weekdays between 08:00–12:30 and 13:45–17:30. These time windows were selected after pilot trials showed significantly lower classification accuracy when neighboring apartments were occupied.

4 METHODOLOGY

This section outlines the full processing and modeling pipeline used to transform raw CSI data into room-level activity predictions. We begin by introducing the design rationale and formal threat model that define the one-sided sensing scenario and its constraints. Next, we describe the preprocessing steps applied to clean and standardize the data, followed by the supervised and clustering-based learning strategies used to infer activity. The goal is to provide a transparent account of the modeling choices and how they align with the study's objective of evaluating generalization across environments and transmitter configurations.

4.1 Design Rationale and Threat Model

The system was designed to reflect a realistic low-resource adversarial setting using only off-the-shelf hardware and minimal computational power. All sensing was performed with ESP32C6 microcontrollers, without relying on access points, internal infrastructure, or

interior placement. This setup mimics conditions where an external observer attempts localization using commodity devices alone.

To match these constraints, all models were chosen for their low computational footprint and ease of deployment. A compact convolutional neural network (CNN) was selected for supervised classification, offering a strong balance between accuracy and efficiency (see Appendix A for architecture details). More complex architectures such as LSTMs or Transformers were evaluated but showed no consistent benefit. Clustering and semi-supervised algorithms were likewise selected based on their simplicity and ability to run without GPU acceleration or large memory requirements. All models, including the CNN and clustering methods, completed training or fitting in under five minutes on a standard desktop CPU (Intel Core i5-7500), suggesting potential for future deployment on embedded or mobile platforms.

4.2 Preprocessing and Feature Extraction

Raw CSI data was first loaded and processed by calculating the amplitude of each complex subcarrier. Null subcarriers were excluded prior to further processing as they carry no data. To reduce noise and artifacts, three sequential filtering steps were applied: a Hampel filter [19] was used to remove outliers, followed by Hamming window smoothing [21] to attenuate short-term fluctuations, and finally wavelet-based denoising [8] to reduce multiscale signal noise.

After filtering, the data was normalized using standard score normalization (z-score) across all samples. The normalized dataset was then segmented by class and subcarrier, and converted into windows of 2 seconds (40 frames at 20 Hz). Each resulting sample had the shape 40 (time steps) × 52 (subcarriers). An illustration of a spectrogram before and after preprocessing is shown in Figure 3.

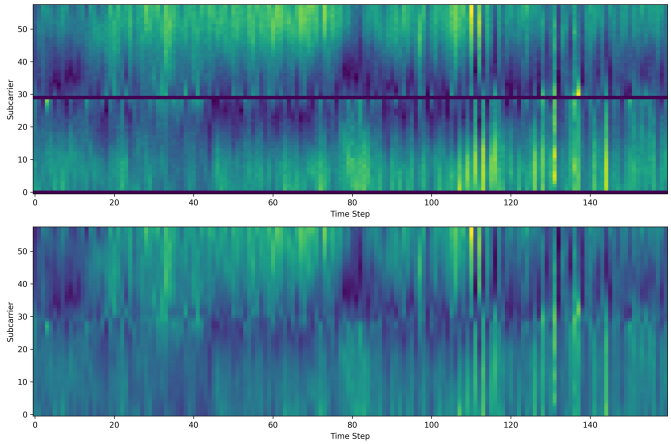


Fig. 3. Example 8 second CSI spectrogram before (top) and after (bottom) preprocessing for kitchen activity in apartment A.

4.3 Supervised Learning

Windowed samples were used to train a convolutional neural network classifier. The full architecture is provided in Appendix A.

Stratified 10-fold cross-validation was performed to evaluate the model, and the entire process was repeated five times to ensure statistical stability. For each run, classification metrics such as accuracy, precision, recall, and F1-score were recorded.

The model was trained using the Adam optimizer with a learning rate of 0.001, a batch size of 16, and for 20 epochs. The loss function was categorical cross-entropy, suitable for multi-class classification.

4.4 Dimensionality Reduction and Unsupervised Clustering

For unsupervised clustering, the filtered and windowed CSI data underwent two-stage dimensionality reduction. First, we applied temporal Principal Component Analysis (tPCA) [10] to each subcarrier, reducing the 40 time steps per subcarrier to 10 temporal components. This compressed representation was then flattened across all subcarriers into a single feature vector for each window. A second PCA step was applied to these vectors to reduce overall dimensionality to 5.

The resulting 5-dimensional vectors were clustered using K-Means and Gaussian Mixture Models. Clustering performance was evaluated using Hungarian-mapped accuracy and F1 score.

4.5 Semi-Supervised Clustering

The same PCA-reduced data was used for semi-supervised clustering using the MPCK-Means-M-F and COP-KMeans algorithms. A 20-fold stratified cross-validation protocol was adopted to simulate a low-label scenario. In each iteration, one fold (approximately 5% of the data) was treated as labeled, while the remaining 19 folds were considered fully unlabeled.

From the labeled fold, pairwise constraints were generated as follows: all sample pairs with the same class label were assigned must-link constraints, and all pairs with different class labels were assigned cannot-link constraints. These constraints were used during clustering, but the labeled samples themselves were excluded from evaluation.

For each run clustering performance was assessed using Hungarian-mapped accuracy and F1 score.

5 RESULTS

This section presents the performance outcomes of the proposed activity recognition approach across several experimental axes. The analysis is structured to reflect the primary evaluation goals: assessing performance in a controlled setting, comparing supervised and clustering models, and examining generalization across environments. Each subsection highlights model behavior, using both accuracy and F1 score metrics (reported as mean ± standard deviation in percentage). Full performance tables, including all accuracy and F1 score values across models and configurations, are provided in Appendix B.

5.1 Performance in a Controlled Setting (Apartment A)

Experiments in Apartment A, a conventional concrete-walled environment, yielded consistently high performance across all configurations and models. Supervised classification using a CNN achieved strong room-level recognition, with accuracy ranging from 91.8% ± 5.5

(Open-Left) to 96.1% ± 4.5 (Boxed-Front), and F1 scores closely matching accuracy across all settings. These results demonstrate the robustness of CNN-based classification in stable indoor environments with limited signal interference.

Clustering-based methods also performed reasonably well in this environment. MPCK-Means-M-F reached up to 80.3% ± 14.7 accuracy and 79.1% ± 15.6 F1 score (Boxed-Front), outperforming other clustering approaches. COP-KMeans trailed behind with accuracy and F1 scores typically in the 55–72% range. GMM clustering showed surprisingly strong results in some cases, reaching up to 80.8% ± 14.6 accuracy. In contrast, K-Means lagged behind, showing both lower mean performance and higher variance.

5.2 Comparison of Supervised and Clustering Models

Across all configurations in Apartment A, the CNN consistently outperformed clustering-based approaches. The average CNN accuracy across all setups was approximately 93.3%, with a comparable F1 score. MPCK-Means-M-F was the top clustering model, followed by GMM. COP-KMeans and K-Means performed less reliably, often falling below 60%.

Clustering methods also exhibited substantially higher standard deviation compared to CNN, indicating greater sensitivity to initialization and environmental variation. Despite these limitations, semi-supervised approaches such as MPCK-Means-M-F provide a viable compromise in settings where fully labeled data is unavailable. Figure 4 summarizes these results visually.

5.3 Generalization to a Different Apartment Layout (Apartment B)

Performance in Apartment B — a less controlled setting with thinner walls, more glass surfaces, and a different spatial layout — was lower and more variable than in Apartment A. The CNN remained the top performer, achieving 80.6% ± 5.5 (Boxed-Front) in its best configuration, but dropping to 43.7% ± 8.9 (Open-Left) in the worst. F1 scores followed the same trend, highlighting the impact of environmental variability on signal propagation. This performance degradation across setups is visualized in Figure 5.

Clustering models were more strongly affected by the domain shift. MPCK-Means-M-F reached a maximum of 62.9% ± 10.9 accuracy and 62.3% ± 10.8 F1 (Boxed-Front), while most other configurations yielded results below 50%. GMM peaked at 57.3% ± 9.7 in Open-Right, but exhibited similar degradation in challenging layouts. COP-KMeans and K-Means offered little robustness, with multiple configurations hovering at or below random baseline levels.

6 DISCUSSION

This section interprets the key findings from the experimental results in light of the study's goals and practical relevance. We examine how model performance was affected by environmental variation, transmitter configuration, and algorithmic approach, with a particular focus on the generalizability across apartments. The discussion also addresses key strengths of the study—such as its one-sided sensing design and use of real apartment layouts—as well as important limitations, including data labeling constraints and unmeasured environmental factors. These reflections inform the practical implications of the work and suggest avenues for future research.

6.1 Impact of shielding and orientation

Shielding effects varied depending on both the environment and the transmitter's orientation. In Apartment A, adding an aluminum-lined box around the transmitter had little or no consistent effect on classification or clustering performance. For example, supervised

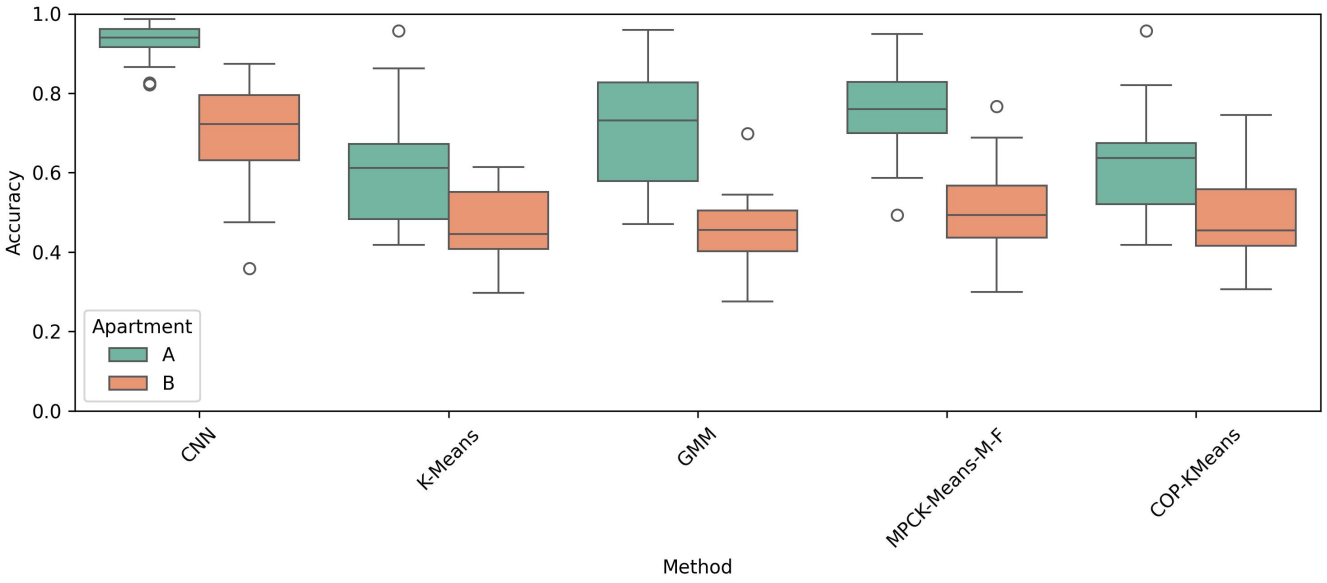


Fig. 4. Classification accuracy across model types for both apartments.

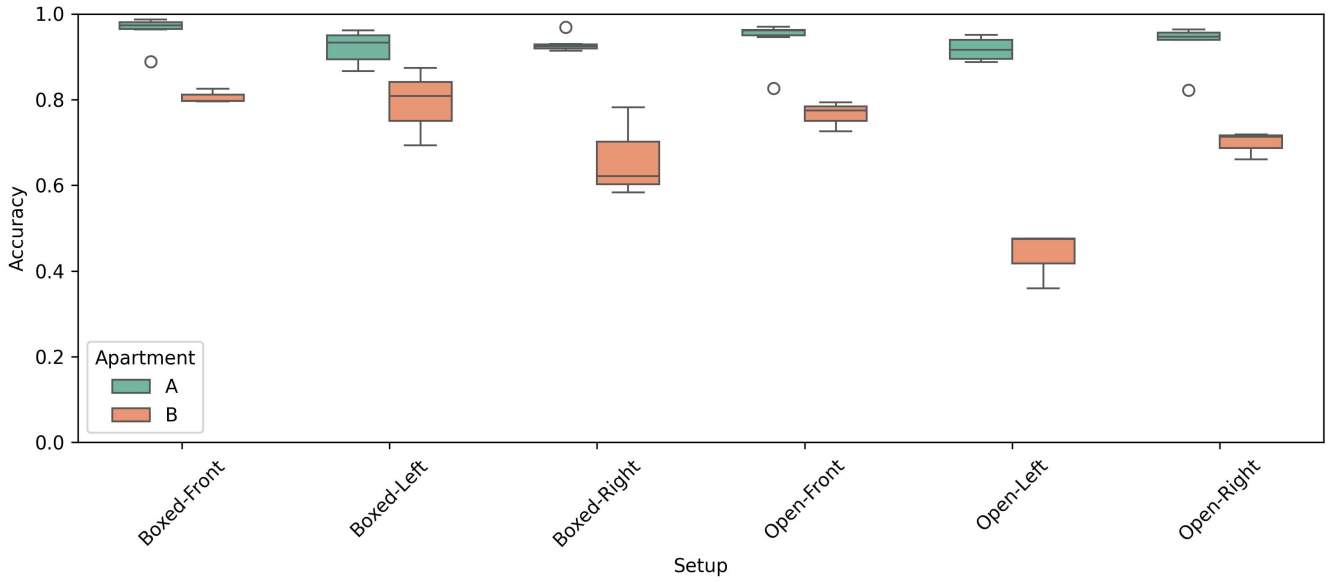


Fig. 5. Comparison of CNN classification accuracy across transmitter configurations in both apartments.

classification achieved 96.1% ± 4.5 accuracy and 96.1% ± 4.5 F1 in the Boxed-Front setup and 93.8% ± 6.2 accuracy and 93.7% ± 6.3 F1 in Open-Front—differences well within the range of expected variation. Similar patterns were observed across all other orientations in Apartment A, suggesting that shielding did not meaningfully alter signal propagation in this more structurally stable environment.

In Apartment B, shielding generally had a limited effect as well—except in one configuration. When the transmitter was oriented left, directly toward the receiver (see Figure 1), performance differed dramatically depending on shielding: Boxed-Left reached 79.2% ± 10.0 accuracy and 78.7% ± 10.2 F1, while Open-Left fell to 43.7% ± 8.9 accuracy and 42.5% ± 9.0 F1—the lowest across all test conditions. This large gap was not observed in other orientations. For example, shielding made little difference between Boxed-Front (80.6% ± 5.5 accuracy, 79.9% ± 6.0 F1) and Open-Front (76.5% ± 6.3 accuracy, 76.2% ± 6.5 F1), or between Boxed-Right (66.2% ± 11.7 accuracy, 65.2% ± 12.4 F1) and Open-Right (69.8% ± 8.1 accuracy, 69.1% ± 8.6 F1).

These findings suggest that shielding may help improve signal consistency in cases where the transmitter is oriented directly at the receiver, especially in reflective environments like Apartment B. In such cases, the enclosure may narrow the beam path or suppress undesired side reflections. Outside of this specific setup, however, shielding did not consistently improve classification performance. This trend is clearly illustrated in Figure 5 and points to a context-specific role for shielding that depends on geometric alignment and environmental complexity.

6.2 Accuracy of HAR using different learning strategies

The supervised convolutional neural network (CNN) consistently achieved the highest performance across both apartments. Averaged over all configurations, it reached a mean accuracy of 83.2% and a mean F1 score of 82.6%, with particularly strong results in Apartment A (e.g., 96.1% ± 4.5 accuracy in Boxed-Front). These results confirm that supervised deep learning methods are highly effective when labeled training data is available, offering both high accuracy and relatively low variance across conditions.

Semi-supervised clustering, particularly MPCK-Means-M-F, showed moderate but consistent performance. Its average accuracy was 62.6%, with a mean F1 score of 61.5%. While it did not match CNN performance, it consistently outperformed unsupervised models across most configurations and showed less degradation when transitioning to Apartment B. This suggests that even limited supervision—in the form of pairwise constraints—can substantially improve clustering outcomes in new environments.

Unsupervised clustering methods performed less reliably overall. GMM was the strongest among them, achieving a mean accuracy of 61.1% and mean F1 score of 58.9%. However, its high variance and inconsistent results across configurations—particularly in Apartment B—limit its practical utility. K-Means and COP-KMeans both averaged below 60% accuracy, with overall F1 scores of 56.4% and 56.8% respectively, and struggled to maintain stability in more complex layouts.

These findings underscore the importance of supervision in CSI-based activity recognition. Supervised CNN models provide the most robust and accurate performance. When labeling is limited, semi-supervised methods such as MPCK-Means-M-F offer a viable compromise. In contrast, unsupervised clustering approaches remain too unstable for reliable deployment without domain adaptation or tuning.

6.3 Generalization across different apartments and participants

Results show a clear and substantial performance drop when moving from Apartment A to Apartment B. In Apartment A, classification and clustering models performed reliably and with low variability across configurations. In Apartment B, however, overall accuracy was lower and much more sensitive to transmitter placement. For example, the CNN model achieved up to 95.7% ± 4.8 in Boxed-Front (A), but only 80.4% ± 7.1 in Boxed-Front (B), and just 41.5% ± 9.3 in Open-Left (B). These results suggest that apartment layout and construction materials—such as thinner walls, metal framing, and increased glass surface area in Apartment B—introduce signal distortions that make CSI-based sensing more difficult.

Participant-specific analysis showed no consistent differences in performance across individuals. All participants contributed data under each configuration, and no systematic advantage or disadvantage was observed (see Table 3). This suggests that in the context of this study, participant movement patterns and body types had a negligible impact on the sensing results.

Overall, these findings indicate that the sensing approach is far more sensitive to environmental structure than to user variation. Future research should further explore how spatial features like wall composition, room geometry, and signal leakage affect sensing robustness, as these are likely to be the limiting factors in real-world deployments.

6.4 Implications for Real-World Deployment

While this study demonstrates that room-level activity inference is technically feasible using external Wi-Fi sensing, the real-world implications are more limited. Even in the best-performing conditions, reliable classification required controlled settings, stable signal paths, and a highly specific apartment structure. In one apartment, semi-supervised clustering reached accuracies that could plausibly compromise a resident's privacy—suggesting that under the right conditions, passive presence detection from outside is possible.

However, replicating those conditions in practice is non-trivial. The sensing setup requires placing visible hardware near the target apartment, including a receiver mounted to an exterior wall and a transmitter positioned near the door. These placements are easy to notice and difficult to conceal, especially in residential settings. Moreover, results varied significantly across environments, indicating that signal propagation—and therefore inference reliability—is highly sensitive to building layout and materials.

In summary, while the potential for privacy intrusion exists, the barriers to practical deployment remain high. One-sided sensing may eventually become more viable as hardware improves, but under current conditions, it is more a proof of concept than an immediate surveillance threat. Further work is needed to assess how generalizable these methods are across diverse settings, and to explore possible mitigations for privacy-aware wireless design.

6.5 Strengths and Limitations

This study presents several notable strengths. First, it explores a novel and largely unexamined sensing configuration: fully external, one-sided through-wall Channel State Information sensing. This design is both technically challenging and relevant to ongoing discussions about wireless privacy. Second, the use of low-cost ESP32-C6 hardware and basic directional antennas reflects a realistic, low-resource adversarial model. Third, the dataset is substantial, consisting of over 600 minutes of labeled CSI recordings across two distinct apartments, with varied orientations, shielding conditions, and participant behaviors. Fourth, the study evaluates multiple machine learning strategies—including supervised, unsupervised, and semi-supervised approaches—providing a broad assessment of inference capabilities under different label availability conditions. Finally, testing was conducted in a disciplined manner, with time-of-day scheduling to reduce ambient interference and randomized activity ordering to minimize bias.

Despite these strengths, the study has important limitations. First, participant activity was not verified using video or motion sensors, so all ground truth labels rely on protocol compliance. Any deviation from instructed behavior could have introduced labeling noise. Second, the environmental diversity was limited: experiments were conducted in only two apartments, which may not capture the variability present in other environments. Third, while recording times were chosen to avoid known peak usage periods, no quantitative measurements of background RF noise or signal-to-noise ratios were collected. As a result, the true interference conditions remain uncharacterized. Fourth, although transmitter shielding was varied (open-air vs. aluminum-lined box), the physical impact of this shielding on signal properties was not directly measured, leaving the observed performance differences difficult to interpret. Fifth, only a single CNN model architecture was evaluated. This layout was selected based on early success on initial datasets, but it is likely that alternative architectures or tuning strategies could yield even higher performance across some configurations. Sixth, for semi-supervised clustering, only a single level of supervision was tested—that is, a fixed number of labeled samples was used to define must-link and cannot-link constraints. It remains possible that using fewer or more labeled samples could significantly affect both clustering accuracy and variability, particularly for MPCK-Means-M-F.

7 CONCLUSION

This paper explored the feasibility of one-sided through-wall human activity recognition using commodity Wi-Fi hardware. By collecting over 600 minutes of CSI data across two residential environments, we evaluated the performance of supervised, unsupervised, and semi-supervised learning models in classifying room-level presence from outside the building. Results showed that high classification accuracy is achievable in structurally favorable settings, even without interior access or line-of-sight.

However, performance varied significantly across apartments, indicating that environmental factors such as wall composition, layout, and signal multipath effects play a critical role in sensing reliability. Among the evaluated methods, supervised CNN models performed best, while semi-supervised clustering showed promise in low-label scenarios, though with higher variance.

Overall, the findings demonstrate that external Wi-Fi sensing can enable spatial inference under the right conditions, but its effectiveness remains highly context-dependent. Future work should focus on improving robustness across diverse environments.

REFERENCES

[1] Espressif Systems 2025. *Espressif/Esp-Csi.* Espressif Systems. https://github.com/espressif/esp-csi

[2] Fadel Adib and Dina Katabi. 2013. See through walls with WiFi!. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*. 75–86. doi:10.1145/2486001.2486039

[3] Alfa Network. 2025. *APA-M25 Dual-Band Directional Panel Antenna.* https://alfa-network.eu/apa-m25

[4] Mikhail Bilenko, Sugato Basu, and Raymond J. Mooney. 2004. Integrating Constraints and Metric Learning in Semi-Supervised Clustering. In *Twenty-First International Conference on Machine Learning - ICML '04* (Banff, Alberta, Canada, 2004). ACM Press, 11. doi:10.1145/1015330.1015360

[5] Hicham Boudlal, Mohammed Serrhini, and Ahmed Tahiri. 2023. A Monitoring System for Elderly People Using WiFi Sensing with Channel State Information. *International Journal of Interactive Mobile Technologies* 17, 12 (2023).

[6] Hui-Hsin Chen, Chi-Lun Lin, and Chun-Hsiang Chang. 2023. WiFi-Based Detection of Human Subtle Motion for Health Applications. 10, 2 (2023), 228. Issue 2. doi:10.3390/bioengineering10020228

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. 39, 1 (1977), 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x

[8] David L Donoho and Iain M Johnstone. 1994. Ideal Spatial Adaptation by Wavelet Shrinkage. 81, 3 (1994), 425–455. doi:10.1093/biomet/81.3.425

[9] Espressif Systems. 2025. *ESP32-C6 Wi-Fi 6 & BLE 5 & Thread/Zigbee SoC | Espressif Systems.* https://www.espressif.com/en/products/socs/esp32-c6

[10] Jiaxin Gao, Wenbo Hu, and Yuntian Chen. 2024. Revisiting PCA for time series reduction in temporal dimension. arXiv:2412.19423 [cs.LG] https://arxiv.org/abs/2412.19423

[11] Francesco Gringoli, Jonas Schulz-Zander, Thomas Marzetta, and Michael Zink. 2019. Free your CSI: A Channel State Information extraction platform for modern Wi-Fi chipsets. *Computer Communications* 133 (2019), 79–90. doi:10.1145/3349623.3355477

[12] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review* 41, 1 (2011), 53–53. doi:10.1145/1925861.1925870

[13] Steven M. Hernandez and Eyuphan Bulut. 2021. Adversarial Occupancy Monitoring Using One-Sided Through-Wall WiFi Sensing. In *ICC 2021 - IEEE International Conference on Communications* (2021-06). 1–6. doi:10.1109/ICC42927.2021.9500267

[14] Steven M. Hernandez and Eyuphan Bulut. 2020. Lightweight and Standalone IoT Based WiFi Sensing for Active Repositioning and Mobility. In *21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM) 2020)* (Cork, Ireland, 2020-06). doi:10.1109/WoWMoM49955.2020.00056

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-Based Learning Applied to Document Recognition. 86, 11 (1998), 2278–2324. doi:10.1109/5.726791

[16] Jian Liu, Hongbo Liu, Yingying Chen, Yan Wang, and Chen Wang. 2019. Wireless sensing for human activity: A survey. *IEEE Communications Surveys & Tutorials* 22, 3 (2019), 1629–1645. doi:10.1109/COMST.2019.2934489

[17] J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. https://api.semanticscholar.org/CorpusID:6278891

[18] Susanna Mosleh, Jason B Coder, Christopher G Scully, Keith Forsyth, and Mohamad Omar Al Kalaa. 2022. Monitoring respiratory motion with Wi-Fi CSI: Characterizing performance and the BreatheSmart algorithm. *IEEE Access* 10 (2022), 131932–131951. doi:10.1109/ACCESS.2022.3230003

[19] Ronald K. Pearson, Yrjö Neuvo, Jaakko Astola, and Moncef Gabbouj. 2016. Generalized Hampel Filters. 2016, 1 (2016), 87. doi:10.1186/s13634-016-0383-6

[20] Random.org. 2025. *RANDOM.ORG - List Randomizer.* https://www.random.org/lists/

[21] Julius O. Smith. 2025. *Spectral Audio Signal Processing.* https://ccrma.stanford.edu/~jos/sasp/ Online book, 2011 edition.

[22] P Sruthi, Sriyanka Satapathy, and Siba K Udgata. 2024. HandFi: WiFi Sensing based hand gesture recognition using channel state information. *Procedia Computer Science* 235 (2024), 426–435. doi:10.1016/j.procs.2024.04.042

[23] Julian Strohmayer and Martin Kampel. 2023. WiFi CSI-Based Long-Range Through-Wall Human Activity Recognition with the ESP32. In *Computer Vision Systems* (Cham, 2023), Henrik I. Christensen, Peter Corke, Renaud Detry, Jean-Baptiste Weibel, and Markus Vincze (Eds.). Springer Nature Switzerland, 41–50. doi:10.1007/978-3-031-44137-0_4

[24] Kiri Wagstaff, Claire Cardie, Seth Rogers, and Stefan Schrödl. 2001. Constrained K-means Clustering with Background Knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning* (San Francisco, CA, USA, 2001-06-28) *(ICML '01)*. Morgan Kaufmann Publishers Inc., 577–584. https://web.cse.msu.edu/~cse802/notes/ConstrainedKmeans.pdf

[25] Dazhuo Wang, Jianfei Yang, Wei Cui, Lihua Xie, and Sumei Sun. 2021. Multimodal CSI-based Human Activity Recognition Using GANs. PP (2021), 1–1. doi:10.1109/JIOT.2021.3080401

[26] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2015-09-07) *(MobiCom '15)*. Association for Computing Machinery, 65–76. doi:10.1145/2789168.2790093

[27] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. 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* (New York, NY, USA, 2014-09-07) *(MobiCom '14)*. Association for Computing Machinery, 617–628. doi:10.1145/2639108.2639143

[28] Yu Xie, Jialin Xiong, and Kyle Jamieson. 2015. *Atheros CSI tool.* Technical Report. Princeton University. https://wands.sg/research/wifi/AtherosCSI/ Technical report.

[29] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon, and Shahrokh Valaee. 2017. A Survey on Behavior Recognition Using WiFi Channel State Information. *IEEE Communications Magazine* 55 (10 2017), 98–104. doi:10.1109/MCOM.2017.1700082

A CNN ARCHITECTURE

The CNN model used for supervised classification consists of two convolutional layers followed by max pooling, a flattening operation, and two fully connected layers. The final layer outputs class probabilities via softmax activation. The detailed architecture is summarized in Table 1.

Table 1. CNN architecture used for CSI-based activity classification

Layer	**Description**
Input	$40 \times 52 \times 1$ CSI window
Conv2D	32 filters, 2×2, ReLU, same padding
MaxPooling2D	3×3
Conv2D	16 filters, 2×2, ReLU, same padding
MaxPooling2D	2×2
Flatten	–
Dense	64 units, ReLU
Dropout	0.1
Dense	Softmax, output = number of classes

AI STATEMENT

Portions of this paper were refactored using OpenAI's ChatGPT-4o to improve clarity, structure, and academic tone. All analysis, results, and interpretations are original, and full responsibility for the content remains with the author.

B DETAILED RESULTS

This section presents the complete classification results across all experimental conditions. Table 2 summarizes the performance of all models—including supervised (CNN), unsupervised (K-Means, GMM), and semi-supervised (MPCK-Means-M-F, COP-KMeans)—across different apartment setups and transmitter configurations. Table 3 provides a detailed breakdown of performance by participant for both apartments, reporting accuracy and F1 scores for each model.

Table 2. Performance comparison of supervised (CNN) and clustering-based (K-Means, GMM, MPCK-Means-M-F, COP-KMeans) methods across apartment setups. All values are mean ± standard deviation in percentage.

Apartment	Setup	CNN		K-Means		GMM		MPCK-Means-M-F		COP-KMeans	
		Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
A	Boxed-Front	96.1 ± 4.5	96.1 ± 4.5	71.6 ± 16.0	69.1 ± 17.2	80.8 ± 14.6	78.3 ± 17.7	80.3 ± 14.7	79.1 ± 15.6	72.2 ± 15.1	70.8 ± 15.8
A	Boxed-Left	92.1 ± 5.9	92.0 ± 6.0	58.0 ± 10.2	55.4 ± 13.9	71.0 ± 14.4	66.4 ± 19.4	73.4 ± 10.2	71.5 ± 12.1	60.3 ± 7.2	59.0 ± 8.4
A	Boxed-Right	93.0 ± 4.6	92.8 ± 4.8	54.0 ± 9.5	48.3 ± 9.7	66.7 ± 12.3	61.7 ± 12.9	77.4 ± 6.6	75.4 ± 7.9	57.5 ± 11.1	53.7 ± 12.5
A	Open-Front	93.8 ± 6.2	93.7 ± 6.3	65.9 ± 12.1	63.4 ± 12.5	80.1 ± 16.7	78.5 ± 18.0	78.6 ± 13.8	78.0 ± 13.8	68.3 ± 11.9	66.9 ± 12.3
A	Open-Left	91.8 ± 5.5	91.7 ± 5.7	54.1 ± 9.0	51.8 ± 7.6	67.7 ± 12.9	62.4 ± 15.3	70.9 ± 9.2	68.9 ± 9.7	55.4 ± 9.6	53.9 ± 9.2
A	Open-Right	92.9 ± 6.0	92.7 ± 6.3	55.4 ± 7.5	51.1 ± 6.7	64.8 ± 14.3	59.1 ± 17.8	72.8 ± 10.1	71.0 ± 11.2	55.3 ± 8.0	52.2 ± 8.2
B	Boxed-Front	80.6 ± 5.5	79.9 ± 6.0	55.9 ± 1.9	54.0 ± 3.5	49.7 ± 2.5	47.6 ± 3.6	62.9 ± 10.9	62.3 ± 10.8	61.2 ± 10.7	60.3 ± 11.3
B	Boxed-Left	79.2 ± 10.0	78.7 ± 10.2	45.3 ± 9.1	46.2 ± 9.8	47.3 ± 8.7	47.8 ± 10.0	50.3 ± 10.8	50.5 ± 11.0	45.6 ± 8.8	46.4 ± 9.6
B	Boxed-Right	66.2 ± 11.7	65.2 ± 12.4	44.1 ± 0.9	44.0 ± 1.6	39.2 ± 5.5	36.9 ± 5.7	48.1 ± 4.9	47.6 ± 5.1	44.7 ± 1.8	44.4 ± 1.7
B	Open-Front	76.5 ± 6.3	76.2 ± 6.5	45.9 ± 6.9	44.7 ± 8.9	43.6 ± 2.4	39.9 ± 5.7	46.7 ± 5.5	45.9 ± 5.7	47.4 ± 6.6	46.9 ± 7.4
B	Open-Left	43.7 ± 8.9	42.5 ± 9.0	36.2 ± 5.7	36.2 ± 5.8	36.5 ± 7.0	35.9 ± 7.0	37.0 ± 5.9	37.0 ± 6.0	37.1 ± 5.5	37.0 ± 5.6
B	Open-Right	69.8 ± 8.1	69.1 ± 8.6	54.7 ± 7.5	52.9 ± 8.5	57.3 ± 9.7	55.3 ± 10.2	57.9 ± 9.6	57.9 ± 9.8	56.1 ± 8.5	56.1 ± 9.4

Table 3. Performance comparison per participant for supervised (CNN) and clustering-based (K-Means, GMM, MPCK-Means-M-F, COP-KMeans) methods across apartments. All values are mean ± standard deviation in percentage.

Participant	Apartment	CNN		K-Means		GMM		MPCK-Means-M-F		COP-KMeans	
		Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
Participant 1	A	92.2 ± 6.0	92.1 ± 6.2	58.6 ± 13.1	55.7 ± 14.1	67.2 ± 15.0	63.3 ± 16.7	73.9 ± 12.1	72.6 ± 12.5	59.6 ± 12.4	57.6 ± 13.2
Participant 2	A	94.7 ± 4.4	94.7 ± 4.5	61.8 ± 13.2	59.0 ± 14.3	75.9 ± 14.6	72.6 ± 17.3	79.2 ± 11.0	78.0 ± 11.9	63.4 ± 13.2	61.7 ± 14.0
Participant 3	A	92.9 ± 6.1	92.7 ± 6.3	59.1 ± 12.1	54.7 ± 13.0	72.4 ± 15.9	67.4 ± 20.6	73.7 ± 10.7	71.3 ± 12.2	61.5 ± 11.8	58.9 ± 12.7
Participant 1	B	64.8 ± 16.6	64.0 ± 17.0	41.4 ± 9.2	40.3 ± 9.6	39.1 ± 7.4	37.1 ± 7.5	45.3 ± 10.0	45.0 ± 10.0	41.6 ± 8.3	41.2 ± 8.5
Participant 2	B	70.6 ± 14.7	69.8 ± 15.2	48.1 ± 7.0	46.8 ± 5.2	49.4 ± 10.2	46.6 ± 10.9	55.3 ± 13.5	54.8 ± 13.5	52.9 ± 12.4	52.8 ± 12.4
Participant 3	B	72.6 ± 13.2	72.0 ± 13.5	51.6 ± 7.6	51.9 ± 8.2	48.3 ± 7.0	47.9 ± 7.9	50.7 ± 9.0	50.8 ± 9.2	51.6 ± 7.6	51.7 ± 8.3