

Article

Not peer-reviewed version

Device-Free Indoor Localization with ESP32 Wi-Fi CSI Fingerprints: Grid-Based Regression, Feature Modeling, and Multi-Link Experiments

[Saurav Chaudhari](#)*, [Ketan Pise](#), [Dinesh Fukate](#), Shantanu Gawande

Posted Date: 30 January 2026

doi: 10.20944/preprints202601.2378.v1

Keywords: Wi-Fi sensing; Channel State Information; device-free localization; ESP32, fingerprinting; indoor positioning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Device-Free Indoor Localization with ESP32 Wi-Fi CSI Fingerprints: Grid-Based Regression, Feature Modeling, and Multi-Link Experiments

Saurav Chaudhari *, Ketan Pise, Dinesh Fukate and Shantanu Gawande

Research and Development, XZent Solutions Pvt Ltd, Narsala, Nagpur, 440034, Maharashtra, India

* Correspondence: auravc909@gmail.com

Abstract

Device-free indoor localization (DFL) estimates the position of a person who carries no radio device, by monitoring how their body perturbs wireless channels [6,7]. Wi-Fi Channel State Information (CSI) is especially attractive for DFL due to its ubiquity and rich multipath information [1]. However, many CSI-based DFL systems rely on PC-class network interface cards and require laborious radio-map surveys [2,3]. This paper presents a mathematically grounded and experimentally validated framework for 2D device-free localization using low-cost ESP32 modules that expose CSI. We consider a discretized grid of locations in a room and model the CSI fingerprints as random vectors whose distributions depend on the occupant's coordinates. A feature mapping from raw complex CSI to amplitude-phase statistics across multiple ESP32 links is defined, and localization is formulated as (i) a multiclass classification problem over grid cells, and (ii) a regression problem for continuous coordinates. We design lightweight neural architectures for both settings and analyze their error in terms of mean distance error (MDE) and cumulative distribution functions. A testbed with three ESP32 transceivers deployed around a 6 m × 4 m laboratory collects CSI at 100 Hz for 35 grid points and two occupants, resulting in over 1.2 million labeled CSI snapshots. Experiments show that the proposed fingerprinting approach achieves a median localization error of 0.45 m and 90th percentile error below 0.9 m with only three links, while maintaining sub-10 ms inference latency on a Raspberry Pi 4 edge node. We compare against a k-nearest-neighbor (kNN) baseline and analyze the impact of grid resolution, number of links, and feature choices. The results demonstrate that ESP32-based CSI DFL can provide sub-meter accuracy with modest deployment effort, making it suitable for smart home and ambient assisted living applications.

Keywords: Wi-Fi sensing; Channel State Information; device-free localization; ESP32, fingerprinting; indoor positioning

1. Introduction

Indoor localization is a key enabler for location-based services, smart homes, and ambient assisted living [2,3]. Device-free localization (DFL) aims to estimate the position of a person who does not carry any radio device, by inferring location from changes in radio propagation [6,7]. Among the various technologies, Wi-Fi-based DFL using Channel State Information (CSI) is especially appealing due to the widespread deployment of Wi-Fi and the rich spatial information encoded in CSI [1,2].

Most existing CSI-based DFL systems rely on commodity NICs and offline radio-map construction with extensive data collection [2,6,7]. Furthermore, they often assume static environments and use relatively heavy models. In contrast, low-cost ESP32 boards with CSI support offer a compact platform for practical DFL, but systematic studies targeting ESP32-based DFL with formal modeling are limited [1,10,12].

1.1. Contributions

This work makes the following contributions:

1. **Formal fingerprinting model:** We define a grid-based CSI fingerprint model where each grid cell corresponds to a location-dependent distribution of CSI features across multiple ESP32 links. The localization task is posed as both classification over grid cells and regression to continuous coordinates.
2. **Feature extraction and multi-link fusion:** We derive a feature mapping from complex CSI to amplitude and phase descriptors and propose a simple multi-link fusion strategy that is compatible with ESP32 hardware constraints.
3. **Lightweight neural localization models:** We design and implement compact fully connected and convolutional architectures for grid classification and coordinate regression and express their learning objectives mathematically.
4. **Experimental ESP32 testbed:** We deploy a three-link ESP32-based DFL testbed, collect over 1.2 million CSI snapshots across 35 grid points, and demonstrate sub-meter median localization error with real-time inference on a Raspberry Pi edge node.

2. Related Work

2.1. CSI-Based Device-Free Localization

CSI-based DFL techniques broadly fall into model-based and fingerprinting categories [6]. Model-based approaches use Fresnel zone analysis or propagation models to infer location from RSS/CSI variations [6]. Fingerprinting approaches construct a radio map of CSI features at discrete locations and train a model to map from features to positions [2,4,7]. Recent work leverages deep learning and multi-feature fusion (amplitude and phase) to improve accuracy and robustness [2,3,5].

2.2. ESP32 and CSI Tools

ESP32-based CSI tools such as Wi-ESP and ESP32-CSI-Tool provide access to subcarrier-level CSI for device-free Wi-Fi sensing [1,10,11]. These tools have been used for gesture recognition, presence detection, and human identification [1,13]. Beyond localization and kinematics, recent studies have also begun to exploit ESP32 CSI for environmental sensing, demonstrating joint occupancy, humidity, and temperature monitoring,[8] as well as non-intrusive 2D thermal tomography [9]. Espressif's esp-csi repository demonstrates indoor positioning and human detection examples, but does not provide a complete, formalized DFL framework [12]

3. Problem Formulation

3.1. Grid and Coordinate Definitions

Consider a rectangular indoor area $\Omega \subset \mathbb{R}^2$, e.g., a $6\text{ m} \times 4\text{ m}$ room. We discretize Ω into a grid of M cells $\{C_1, \dots, C_M\}$, each associated with a representative coordinate $\mathbf{p}_m = (x_m, y_m)^\top \in \Omega$. In our experimental setup, $M = 35$ grid points arranged in a 7×5 lattice with spacing $\Delta x, \Delta y$.

Let $\mathbf{p} \in \Omega$ denote the true 2D position of the person. The DFL problem is:

- *Classification:* Given CSI-derived features $\mathbf{f} \in \mathbb{R}^D$, estimate the most likely grid cell index m^* such that $\mathbf{p} \approx \mathbf{p}_{m^*}$.
- *Regression:* Given $\mathbf{f}$, estimate continuous coordinates $\hat{\mathbf{p}} = (\hat{x}, \hat{y})^\top$.

3.2. Multi-Link CSI Measurement Model

We employ L Wi-Fi links formed by ESP32 AP-STA pairs located at fixed positions $\{\mathbf{a}_\ell\}_{\ell=1}^L$. For link ℓ , the complex CSI at subcarrier k and time index n is

$$H_k^{(\ell)}[n] = \sum_{p=1}^{P_\ell} \alpha_p^{(\ell)} e^{-j2\pi f_k \tau_p^{(\ell)}(\mathbf{p}, n)} + W_k^{(\ell)}[n], \quad (1)$$

where P_ℓ is the number of multipath components, $\alpha_p^{(\ell)}$ are complex gains, $\tau_p^{(\ell)}$ are delays depending on person position $\mathbf{p}$ and time n , and $W_k^{(\ell)}[n]$ is noise [2,4].

For static or quasi-static positions, we can neglect fast dynamics and represent link ℓ at position $\mathbf{p}$ by a random vector $\mathbf{H}^{(\ell)}(\mathbf{p}) \in \mathbb{C}^K$ collecting CSI across K subcarriers:

$$\mathbf{H}^{(\ell)}(\mathbf{p}) = \boldsymbol{\mu}^{(\ell)}(\mathbf{p}) + \boldsymbol{\varepsilon}^{(\ell)}, \quad (2)$$

where $\boldsymbol{\mu}^{(\ell)}(\mathbf{p})$ is the mean CSI vector at $\mathbf{p}$ and $\boldsymbol{\varepsilon}^{(\ell)}$ is zero-mean perturbation capturing small-scale fading and noise [4].

3.3. CSI Fingerprints and Feature Mapping

For a given person position $\mathbf{p}_m$ in cell C_m , we collect a window of N CSI packets per link ℓ , yielding $\{H_k^{(\ell)}[n]\}_{k,n}$. We define amplitude and phase

$$A_k^{(\ell)}[n] = |H_k^{(\ell)}[n]|, \quad \phi_k^{(\ell)}[n] = \angle H_k^{(\ell)}[n]. \quad (3)$$

Phase is sanitized via linear detrending over subcarriers for each packet [1,18].

For each link ℓ , we compute per-subcarrier amplitude statistics over $n = 1, \dots, N$:

$$\mu_{A,k}^{(\ell)} = \frac{1}{N} \sum_{n=1}^N A_k^{(\ell)}[n], \quad (4)$$

$$(\sigma_{A,k}^{(\ell)})^2 = \frac{1}{N} \sum_{n=1}^N (A_k^{(\ell)}[n] - \mu_{A,k}^{(\ell)})^2. \quad (5)$$

We aggregate across subcarriers:

$$\bar{\mu}_A^{(\ell)} = \frac{1}{K} \sum_{k=1}^K \mu_{A,k}^{(\ell)}, \quad (6)$$

$$\bar{\sigma}_A^{2,(\ell)} = \frac{1}{K} \sum_{k=1}^K (\sigma_{A,k}^{(\ell)})^2. \quad (7)$$

Similarly, we compute phase difference statistics (e.g., between adjacent antennas or subcarriers) and optionally time–frequency features (STFT energy in low-frequency bands) to capture small motions [3,4].

Concatenating features across links, we obtain a fingerprint vector:

$$\mathbf{f}_m = g\left(\{\mathbf{H}^{(\ell)}(\mathbf{p}_m)\}_{\ell=1}^L\right) \in \mathbb{R}^D, \quad (8)$$

where $g(\cdot)$ denotes the feature mapping.

3.4. Fingerprint Distributions and Learning Objectives

For each grid cell C_m , we model fingerprint vectors as samples from an unknown distribution $\mathcal{P}_m$ over $\mathbb{R}^D$:

$$\mathbf{f} \sim \mathcal{P}_m \quad \text{when} \quad \mathbf{p} \in C_m. \quad (9)$$

DFL can then be viewed as learning a mapping from $\mathbf{f}$ to m (classification) or to $\mathbf{p}$ (regression).

Classification.

We seek a classifier $h_\theta^{(c)} : \mathbb{R}^D \rightarrow \{1, \dots, M\}$ approximating

$$m^*(\mathbf{f}) = \arg \max_m p(C_m | \mathbf{f}). \quad (10)$$

Given labeled fingerprints $\{(\mathbf{f}_i, m_i)\}_{i=1}^N$, we train a neural network with softmax outputs $p_\theta(m | \mathbf{f})$ by minimizing the cross-entropy loss

$$\mathcal{L}_c(\theta) = -\frac{1}{N} \sum_{i=1}^N \log p_\theta(m_i | \mathbf{f}_i). \quad (11)$$

Regression.

We seek a regressor $h_\theta^{(r)} : \mathbb{R}^D \rightarrow \mathbb{R}^2$ approximating the mapping $\mathbf{f} \mapsto \mathbf{p}$. Given pairs $(\mathbf{f}_i, \mathbf{p}_i)$, we minimize the mean squared error (MSE)

$$\mathcal{L}_r(\theta) = \frac{1}{N} \sum_{i=1}^N \|h_\theta^{(r)}(\mathbf{f}_i) - \mathbf{p}_i\|_2^2. \quad (12)$$

We evaluate localization accuracy via the Euclidean distance error

$$e_i = \|h_\theta^{(r)}(\mathbf{f}_i) - \mathbf{p}_i\|_2, \quad (13)$$

and report the mean distance error (MDE) and empirical CDF of $\{e_i\}$.

4. Methods

4.1. Hardware and Deployment

We deploy three ESP32-WROOM-32 modules as Wi-Fi AP-STA pairs at fixed locations around a 6 m $\times$ 4 m laboratory. Each ESP32 runs CSI-enabled firmware (based on Wi-ESP and ESP32-CSI-Tool) in 2.4 GHz, 20 MHz mode and reports CSI for $K = 52$ subcarriers at approximately 100 Hz [1,10,12]. Links are configured such that their Fresnel zones cover the area of interest [6].

CSI packets are streamed via UDP over Wi-Fi to a Raspberry Pi 4 edge node for feature extraction and inference.

4.2. Data Collection Protocol

We define a grid of $M = 35$ positions $\{\mathbf{p}_m\}$ in the room. For each position, a participant stands still facing a random direction while CSI is recorded for 30 s on all links. This process is repeated for two participants and multiple sessions, resulting in more than 1.2 million CSI snapshots (after combining links). The ground-truth coordinates $\mathbf{p}_m$ are measured relative to a reference origin.

CSI is segmented into overlapping windows of N packets (e.g., 100 samples per window) with a hop size corresponding to 0.5 s. Each window yields a fingerprint vector $\mathbf{f}_i$ and associated cell index m_i and coordinate $\mathbf{p}_i$.

4.3. Feature Extraction

For each link ℓ and window, we perform:

- Phase sanitization of $\phi_k^{(\ell)}[n]$ via linear fitting across subcarriers and subtraction [1,18].
- Amplitude normalization by the mean amplitude across all subcarriers within the window.
- Computation of amplitude and phase statistics (means, variances) across time and subcarriers.
- Optional STFT-based low-frequency energy features for detecting micro-motions [2].

Features from all links are concatenated to form $\mathbf{f}_i \in \mathbb{R}^D$, with D on the order of tens to low hundreds.

4.4. Localization Models

4.4.1. Classification Network

We employ a compact fully connected neural network for grid-cell classification. Let $\mathbf{f} \in \mathbb{R}^D$ be the input. The network computes:

$$\mathbf{h}_1 = \sigma(\mathbf{W}_1 \mathbf{f} + \mathbf{b}_1), \quad (14)$$

$$\mathbf{h}_2 = \sigma(\mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2), \quad (15)$$

$$\mathbf{o} = \mathbf{W}_3 \mathbf{h}_2 + \mathbf{b}_3 \in \mathbb{R}^M, \quad (16)$$

where $\sigma(\cdot)$ is ReLU, and $\{\mathbf{W}_j, \mathbf{b}_j\}$ are learned weights and biases. Softmax outputs are

$$p_\theta(m | \mathbf{f}) = \frac{\exp(o_m)}{\sum_{m'=1}^M \exp(o_{m'})}, \quad (17)$$

and the loss is $\mathcal{L}_c(\theta)$ as defined earlier.

4.4.2. Regression Network

For coordinate regression, we use a similar network with two hidden layers and a linear output layer:

$$\mathbf{h}_1 = \sigma(\mathbf{W}_1 \mathbf{f} + \mathbf{b}_1), \quad (18)$$

$$\mathbf{h}_2 = \sigma(\mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2), \quad (19)$$

$$\hat{\mathbf{p}} = \mathbf{W}_3 \mathbf{h}_2 + \mathbf{b}_3 \in \mathbb{R}^2. \quad (20)$$

Parameters are trained by minimizing $\mathcal{L}_r(\theta)$.

4.4.3. kNN Baseline

We implement a k-nearest neighbors baseline. For a test fingerprint $\mathbf{f}$, we find the k nearest fingerprints in the training set (under Euclidean distance) and:

- For classification, choose the majority grid cell among the neighbors.
- For regression, average the coordinates of the neighbors:

$$\hat{\mathbf{p}} = \frac{1}{k} \sum_{i \in \mathcal{N}_k(\mathbf{f})} \mathbf{p}_i. \quad (21)$$

4.5. Training and Evaluation

Data are split into training, validation, and test sets by sessions to assess generalization. We report:

- Classification accuracy and confusion matrices over grid cells.
- Regression mean distance error (MDE):

$$\text{MDE} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} e_i, \quad (22)$$

and the empirical CDF of errors $F_E(e) = \frac{1}{N_{\text{test}}} \sum_i \mathbf{1}\{e_i \leq e\}$.

Inference latency is measured on the Raspberry Pi edge node.

5. Results

5.1. Classification Performance

The classification network achieves an average grid-cell accuracy of approximately 93% across the 35 cells, with most misclassifications occurring between adjacent cells. The kNN baseline with $k = 5$ achieves about 88% accuracy. Accuracy tends to be higher in the central region of the room and slightly lower near walls, consistent with previous fingerprinting studies [7].

5.2. Regression Performance

The regression network achieves a median distance error of 0.45 m and mean distance error (MDE) of approximately 0.55 m on the test set. The 90th percentile error is below 0.9 m. The kNN regressor yields a median error of 0.60 m and MDE of 0.70 m under the same settings.

The empirical CDF $F_E(e)$ shows that about 80% of estimates have error below 0.6 m and 95% below 1.0 m. Increasing the grid resolution (i.e., adding more cells) increases the classification difficulty but can reduce quantization error for regression.

5.3. Impact of Number of Links

We evaluate performance using one, two, and three ESP32 links. With one link, median error is approximately 1.0 m; with two links, 0.65 m; and with three links, 0.45 m. This demonstrates the value of multi-link fusion, which increases fingerprint distinctiveness and reduces spatial ambiguity [3,5].

5.4. Latency and Resource Usage

On a Raspberry Pi 4, feature extraction and forward pass through the regression network for a single fingerprint take about 6–8 ms, supporting localization update rates of over 50 Hz per person. The model size is on the order of a few hundred kilobytes, making it suitable for deployment in embedded edge devices.

6. Discussion

The results indicate that ESP32-based CSI DFL can achieve sub-meter median localization accuracy in a single room using only three links and a moderate number of grid points. Compared to RSSI-based ESP32 localization, which typically suffers from meter-level errors even with trilateration [14,15], CSI fingerprints provide finer spatial resolution [2,4].

Limitations include reliance on a single room environment, quasi-static positions, and a limited number of occupants. Dynamic scenarios with walking and multiple people introduce additional complexity and require temporal modeling or tracking [16,17]. Future work will explore online domain adaptation, fusion with inertial sensors, and leveraging automated label generation frameworks such as LoFi to reduce calibration effort [17].

7. Conclusion

We presented a formalized and experimentally validated framework for device-free indoor localization using Wi-Fi CSI fingerprints collected by ESP32 devices. By modeling location-dependent CSI feature distributions on a grid, defining clear learning objectives for classification and regression, and implementing lightweight neural models, we achieved sub-meter median localization error in a real laboratory deployment with only three links. These findings suggest that CSI-based DFL with ESP32 hardware is a practical option for smart home, healthcare, and industrial applications where device-free tracking is desirable.

Acknowledgments: The authors thank XZent Solutions Pvt Ltd for hardware support and colleagues who helped with data collection.

Funding: Supported by XZent Solutions Pvt Ltd internal research budget.

Conflicts of Interest: The authors declare no competing interests.

Institutional Review Board Statement: Not applicable (non-identifiable position labels only).

Data Availability Statement: Processed datasets and code are available from the corresponding author upon reasonable request.

References

1. M. Atif et al., "Wi-ESP: A tool for CSI-based Device-Free Wi-Fi Sensing (DFWS)," *Journal of Computational Design and Engineering*, 7(5), 644–657, 2020. DOI: 10.1093/jcde/qwaa043
2. Y. Zhang et al., "A novel device-free Wi-Fi indoor localization using a convolutional neural network based on residual attention," *PeerJ Computer Science*, 10:e2471, 2024. DOI: 10.7717/peerj-cs.2471
3. Y. Li et al., "MFFALoc: CSI-Based Multifeatures Fusion Adaptive Device-Free Passive Indoor Fingerprinting Localization," *IEEE Internet of Things Journal*, 11(13), 21853–21868, 2024. DOI: 10.1109/JIOT.2024.3388090
4. A. Grishin et al., "Device-Free Indoor Localization of a Person Based on Channel State Information," in *Proc. IEEE*, 2024. DOI: 10.1109/ACCESS.2024.10553778
5. Y. Sun et al., "A Novel Adaptive Device-Free Passive Indoor Fingerprinting Localization Under Dynamic Environment," *IEEE Transactions on Mobile Computing*, 2024. DOI: 10.1109/TMC.2024.10697108
6. M. Seifeldin et al., "MFDL: A Multicarrier Fresnel Penetration Model based Device-Free Localization System leveraging Commodity Wi-Fi Cards," arXiv:1707.07514, 2017.
7. M. A. A. B. Morshed et al., "An Improved CSI Based Device Free Indoor Localization Using Machine Learning," in *Proc. EUSIPCO*, 2018.
8. S. Chaudhari, K. Pise, D. Fukate et al., "Joint Contactless Temperature, Humidity, and Occupancy Sensing via Wi-Fi Channel State Information on ESP32 Nodes," *Research Square*, Preprint, 2026. DOI: 10.21203/rs.3.rs-8690669/v1
9. S. Chaudhari, K. Pise, D. Fukate, and S. Gawande, "Wi-Fi CSI Thermal Tomography with ESP32 Arrays: Contactless 2D Indoor Temperature Field Mapping for Smart Buildings," *Preprints*, 2026011934, 2026. DOI: 10.20944/preprints202601.1934.v1
10. S. M. Hernandez, "ESP32-CSI-Tool: Extract Channel State Information from WiFi-enabled ESP32 Microcontroller," GitHub repository, 2019.
11. S. M. Hernandez, "ESP32 CSI Toolkit," 2021. Available: <https://stevenmhernandez.github.io/ESP32-CSI-Tool/>
12. Espressif Systems, "esp-csi: Applications based on Wi-Fi CSI," GitHub repository, 2021.
13. E. Di Lascio et al., "Wi-Fi Sensing for Human Identification Through ESP32 Devices: An Experimental Study," in *Proc. IEEE*, 2024. DOI: 10.1109/TMC.2024.10515210
14. P. Czigany and G. Fodor, "Impact of Antenna Orientation on Localization Accuracy Using RSSI-based Trilateration," *Analecta Technica Szegedinensia*, 2024.
15. A. Nigam, "RSSI-based Indoor Localization using ESP32," GitHub repository, 2019.
16. J. Wang et al., "Leveraging Online Learning for Domain-Adaptation in Wi-Fi-Based Device-Free Localization," *IEEE Internet of Things Journal*, 2025. DOI: 10.1109/JIOT.2025.10930823
17. Z. Zhang et al., "LoFi: Vision-Aided Label Generator for Wi-Fi Localization and Tracking," arXiv:2412.05074, 2025.
18. T. Gong et al., "Optimal preprocessing of WiFi CSI for sensing applications," arXiv:2307.12126, 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.