

Geometry over Depth: Multi-Node WiFi CSI Fusion for Robust Human Activity Recognition

Mojtaba Mansouri Nejad*, Azin Moradbeikie[†], David Beserra[‡], Ahmad Keshavarz*, and Hojat Ghimatgar*

*Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran

[†]Università degli Studi di Trieste, Trieste, Italy

[‡]École Pour l'Informatique et les Techniques Avancées (EPITA), Le Kremlin-Bicêtre, France

M.mansourinejad@mehr.pgu.ac.ir*, azin.moradbeikie@epita.fr[†], david.beserra@epita.fr[‡],

a.keshavarz@pgu.ac.ir*, ghimatgar@pgu.ac.ir*

Abstract—WiFi Channel State Information (CSI) enables contactless, device-free Human Activity Recognition (HAR), but a single transmitter–receiver link observes motion from one fixed geometry and is sensitive to subject position. The common response is to absorb this fragility within increasingly deep models, at the cost of data requirements, interpretability, and a computational footprint ill-suited to low-power nodes. This paper takes a different position that the fragility of single-viewpoint sensing is better addressed through geometry than through model capacity. Using low-cost ESP32 nodes, we test whether spatial diversity from multiple observation points can substitute for model complexity in CSI-based HAR. We extract a set of 43 per-subcarrier features across statistical, nonlinear, communication-oriented, and spectral families, apply a two-stage subcarrier-then-feature selection scheme, and fuse the selected features of two spatially separated nodes for classification with conventional machine learning. Evaluation uses data from three subjects under a subject-activity-aware protocol that prevents train–test leakage, with results aggregated over ten randomized repetitions. Feature-level fusion of two nodes improves recognition accuracy from 75.56% for the best single-node configuration to 80%, providing initial evidence that interpretable fusion of spatially diverse low-cost nodes yields more robust HAR than either node alone under the tested conditions.

Index Terms—WiFi sensing, channel state information, human activity recognition, spatial diversity.

I. INTRODUCTION

Human Activity Recognition (HAR) underpins a growing range of applications in smart homes, ambient assisted living, security, and industrial monitoring [11]. While camera- and wearable-based approaches dominate, both carry well-known drawbacks (vision systems fail in poor lighting or through occlusions, and wearables depend on user compliance) [10], [11]. WiFi Channel State Information (CSI) sensing has emerged as a compelling alternative (contactless, device-free, and illumination-independent) by exploiting the perturbations induced by human motion on the wireless channel [12], [24]. Because it captures no visual imagery, it is also generally regarded as more privacy-preserving than camera-based sensing.

The central difficulty in CSI-based HAR is not capturing the signal but contending with its fragility [19]. A single transmitter–receiver link observes motion from one fixed geometry, so the same activity produces markedly different channel perturbations depending on where and how the subject moves relative to that link.

The common response to this problem has been to use more complex deep learning models [23]. For example, recent work uses attention-based GRU networks combined with pruning and data augmentation to keep large recurrent models small enough to deploy; signal-decomposition pipelines that separate human-related signals before fusing multiple streams [17]; and CNNs that process each CSI channel separately and combine their decisions [18]. These methods improve accuracy, but they all try to solve the problem by making the model larger and deeper. This comes at a cost: they require large amounts of training data, are hard to interpret, and need more computation than the low-power nodes used for commodity CSI sensing can comfortably provide [2], [9]. In effect, the problem of an unreliable single viewpoint is treated as something to be fixed by adding model depth.

This paper takes a different approach. Instead of solving the single-viewpoint problem with a more complex model, we address it through the placement of the sensing hardware. The ESP32 makes CSI measurement available on low-cost microcontrollers, so deploying several nodes is inexpensive and practical. Multi-node and low-power ESP32 datasets have recently been released, confirming that such deployments are feasible [1]–[3]. However, these datasets are used with deep learning models, and it is not yet known whether the spatial diversity provided by multiple nodes can replace model complexity rather than only add to it. We hypothesize that two low-cost nodes observing the same activity from different positions capture complementary information, and that fusing this information allows simple, interpretable classical classifiers to reach the recognition accuracy that would otherwise require deep models.

Studying this substitution is useful because, if spatial diversity can replace model complexity, robust HAR becomes possible without the data and computational cost of deep models, using hardware cheap enough to deploy at scale. We therefore measure the gain in recognition accuracy obtained by fusing multiple low-cost nodes, compared to single-node sensing, using only interpretable classical classifiers. The study is conducted on a CSI dataset collected in our laboratory under controlled conditions, comparing two single-node baselines against their fusion across five conventional classifiers, under a subject-activity-aware protocol that prevents train–test leakage.

To keep the fused representation compact and interpretable, we rank subcarriers and then features in two stages, and to ensure the measured gains are real rather than an artifact of data leakage, we split the data before balancing it.

The primary contribution of this work is the idea, and its validation, that spatial diversity from multiple low-cost ESP32 nodes can replace model complexity in CSI-based HAR, and interpretable feature-level fusion of two nodes yields more robust recognition than either node alone. This central claim is supported by four further contributions:

- A multi-node HAR framework that places low-cost ESP32 nodes at different geometric positions and fuses their features to improve accuracy without deep models.
- A two-stage feature selection scheme that first ranks subcarriers and then features within the retained subcarriers, keeping the fused representation low-dimensional and interpretable.
- A subject-activity-aware evaluation protocol in which training and test data are split before class balancing.
- A controlled CSI dataset, collected in our laboratory using multiple ESP32 nodes, for evaluating spatial diversity in HAR.

The remainder of the paper is organized as follows. Section II reviews related work in CSI-based HAR. Section III details the proposed method. Section IV presents the experimental setup and results, and Section V concludes.

II. RELATED WORK

Early CSI-based HAR relied on specialized WiFi hardware, such as the Intel 5300 NIC with the Linux CSI Tool [7], the Atheros CSI Tool [22], and the Nexmon CSI extractor [4], which limited large-scale and low-cost deployment. Foundational work established how CSI relates to human motion, from early device-free activity identification using CSI signatures [21] to quantitative models linking CSI dynamics to movement speed and activity [20]. The CARM system introduced the CSI-speed and CSI-activity models and reported recognition accuracy of about 96% using classical processing and an HMM classifier [20]. More recently, low-cost ESP32 devices have made CSI measurement available on microcontrollers without a host computer [8], enabling dense and inexpensive sensing for tasks such as crowd counting [3], on-device human activity recognition [2], and the monitoring of robot-assisted therapy [1]. Our work uses the same class of low-cost hardware but focuses on multi-node activity recognition.

The sensitivity of CSI to user position and orientation has motivated the use of spatial diversity [6]. Wang *et al.* proposed WiSDAR, which forms multiple virtual links from the several antennas of a single MIMO device and applies a deep learning model, raising accuracy from about 75% to about 96% and confirming that spatial diversity reduces position-dependent errors [19]. Multi-link fusion was also reported to improve accuracy across environments in the CARM system [20]. Fusion of multiple receivers has likewise been studied for in-vehicle gesture recognition, where a Transformer LSTM

model combined two WiFi receivers [25]. The benefit of collaborative wireless observations has also been demonstrated in indoor positioning, where measurements obtained across distributed infrastructure can improve robustness in dynamic environments [5], [14], [15]. These approaches obtain spatial diversity either from the multiple antennas of a single device or through deep fusion models. In contrast, we obtain spatial diversity from physically separate low-cost ESP32 nodes and combine them at the feature level, without a deep model.

Most recent CSI-HAR systems use deep learning. Hybrid CNN-GRU networks with attention reach very high accuracy on benchmark datasets [13], and lightweight models reduce this cost through pruning and augmentation while keeping a deep backbone [9]. Deep models, however, require large training sets and more computation than low-power nodes can easily provide. Classical machine learning remains attractive on such hardware: in the Wi-CaL system, a LightGBM classifier matched deep learning performance for ESP32-based sensing [3]. Edge-AI has also been investigated in other resource-constrained wireless-sensing applications, demonstrating that locally executed machine-learning models can improve the robustness of radio-signal analysis under environmental variation [16]. We follow this classical direction and study whether spatial diversity from multiple nodes can provide the accuracy that is otherwise sought through deeper models.

III. PROPOSED METHOD

The CSI measurement collected from node n is denoted $\mathbf{X}^{(n)} \in \mathbb{C}^{S \times T}$, where S is the number of subcarriers and T the number of time samples per segment. Given a set of spatially distributed nodes $\mathcal{N}$, the objective is to map the per-node CSI representation to an activity label $y \in \{1, \dots, C\}$, where C is the number of activity classes. In this work, $\mathcal{N} = \{\text{Node 0}, \text{Node 3}\}$.

The proposed pipeline comprises five stages: signal preprocessing, feature extraction, two-stage feature selection, multi-node feature fusion, and classification. Each stage is detailed below.

A. Signal Preprocessing

The CSI amplitude, obtained as the magnitude of the complex CSI measurements, is adopted as the signal representation, as it is more stable than the raw phase, which is corrupted by carrier frequency offset and sampling time offset in commodity hardware. In the preprocessing step, the edge subcarriers 0–4 and 60–63 are discarded, since they correspond to guard-band and null subcarriers that carry no reliable channel information and contribute mainly noise. The remaining subcarriers are retained for feature extraction.

B. Feature Extraction

From each retained subcarrier, a set of 43 features is extracted to characterize the amplitude variations induced by human motion. Let $x = \{x_t\}_{t=1}^T$ denote the amplitude sequence of a subcarrier over a segment of T samples, with mean μ and standard deviation σ . The features are organized into

four families: statistical, nonlinear, communication-oriented, and complementary spectral and difference descriptors. Representative definitions are given below, and the complete set is summarized in Table I.

1) *Statistical features*: Fifteen statistical features describe the amplitude distribution, including the mean μ , standard deviation σ , variance σ^2 , median, minimum, maximum, range, the 25th and 75th percentiles (Q_{25} , Q_{75}), interquartile range ($IQR = Q_{75} - Q_{25}$), mean absolute deviation, skewness, kurtosis, root mean square, and coefficient of variation $CV = \sigma/\mu$, have been considered. The first four moments are defined as Equation 1 and 2.

$$\mu = \frac{1}{T} \sum_{t=1}^T x_t, \quad \sigma^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \mu)^2, \quad (1)$$

$$\gamma_1 = \frac{1}{T} \sum_{t=1}^T \left(\frac{x_t - \mu}{\sigma} \right)^3, \quad \gamma_2 = \frac{1}{T} \sum_{t=1}^T \left(\frac{x_t - \mu}{\sigma} \right)^4, \quad (2)$$

where the skewness γ_1 captures the directional asymmetry of motion and the kurtosis γ_2 the tail-heaviness associated with short, sudden bursts.

2) *Nonlinear features*: Seven nonlinear features describe signal complexity and dynamics. The histogram entropy quantifies the spread of the amplitude distribution,

$$H = - \sum_i p_i \log p_i, \quad (3)$$

where p_i is the probability of amplitude bin i , whereas the permutation entropy measures the ordinal irregularity of the sequence. The three Hjorth parameters describe the signal through its variance and derivatives,

$$\text{Activity} = \sigma_{x'}^2, \quad \text{Mobility} = \frac{\sigma_{x'}}{\sigma_x}, \quad \text{Complexity} = \frac{\text{Mobility}(x')}{\text{Mobility}(x)}, \quad (4)$$

where x' is the first temporal difference of x . Finally, the Petrosian fractal dimension characterizes geometric roughness from the number of sign changes in x' , and the turning-point rate gives the proportion of local extrema, which rises for periodic activities.

3) *Communication-oriented features*: Twelve features adopt a telecommunication perspective on the channel. These include the mean power and total energy,

$$\bar{P} = \frac{1}{T} \sum_{t=1}^T x_t^2, \quad E = \sum_{t=1}^T x_t^2, \quad (5)$$

the RMS gain, and the peak-to-average power ratio,

$$\text{PAPR} = \frac{\max_t x_t^2}{\bar{P}}. \quad (6)$$

An SNR-like stability index $\bar{P}/\sigma^2$ and a stability ratio μ/σ quantify amplitude steadiness, while two fading-range descriptors, $P_{95} - P_5$ and $P_{90} - P_{10}$, characterize the fading envelope robustly to outliers. The fade depth $\mu - \min_t x_t$ captures obstruction-induced signal drops, and the temporal variation gives the mean absolute sample-to-sample change. Short-term temporal coherence is measured by the lag-1 and lag-2 autocorrelation coefficients,

$$\rho_\tau = \frac{\sum_{t=1}^{T-\tau} (x_t - \mu)(x_{t+\tau} - \mu)}{\sum_{t=1}^T (x_t - \mu)^2}, \quad \tau \in \{1, 2\}. \quad (7)$$

TABLE I
COMPLETE SET OF 43 PER-SUBCARRIER FEATURES.

Family	Features
Statistical (15)	mean, std, variance, median, min, max, range, Q_{25} , Q_{75} , IQR, mean absolute deviation, skewness, kurtosis, RMS, coefficient of variation
Nonlinear (7)	permutation entropy, histogram entropy, Hjorth activity, Hjorth mobility, Hjorth complexity, Petrosian fractal dimension, turning-point rate
Communication (12)	mean power, total energy, RMS gain, PAPR, SNR-like index, fading range ($P_{95}-P_5$), fading range ($P_{90}-P_{10}$), fade depth, stability (μ/σ), temporal variation, autocorrelation lag-1, autocorrelation lag-2
Complementary (9)	difference std, difference energy, slope, peak rate, spectral centroid, spectral spread, spectral entropy, dominant frequency index, low-to-high energy ratio

4) *Complementary features*: Nine complementary features capture difference- and frequency-domain behavior. The difference standard deviation and difference energy summarize the dispersion and intensity of the first difference x' ; the overall slope gives the net linear trend across the window; and the peak rate counts local maxima. The remaining five are spectral descriptors derived from the power spectrum $P_k = |X_k|^2$, where $X_k = \sum_{t=1}^T x_t e^{-j2\pi kt/T}$. With normalized spectral power $\hat{p}_k = P_k / \sum_j P_j$, the spectral centroid and spectral spread are computed by using Equation 8.

$$c = \sum_k f_k \hat{p}_k, \quad s = \sqrt{\sum_k (f_k - c)^2 \hat{p}_k}, \quad (8)$$

the spectral entropy is $H_f = - \sum_k \hat{p}_k \log \hat{p}_k$, the dominant frequency index is $\arg \max_k P_k$ with the DC component removed, and the low-to-high energy ratio compares power in the low- and high-frequency bands. The descriptors from all four families are concatenated across the retained subcarriers to form a per-node feature vector.

C. Two-Stage Feature Selection

To prevent information leakage, feature ranking and selection are performed exclusively on the training partition and are subsequently applied unchanged to the test data. Selection proceeds in two stages.

a) *Stage 1 (Subcarrier selection)*: Subcarrier importance is estimated using an Extra-Trees classifier. The subcarriers are ranked by their importance scores, and the top-ranked subset is retained.

b) *Stage 2 (Feature selection)*: Within the reduced subcarrier space, feature importance is estimated using the same Extra-Trees classifier. Individual features are ranked by their importance scores, and the top-ranked subset is retained.

The number of retained subcarriers and features is determined through validation on the training set.

D. Multi-Node Feature Fusion

Stages III-A–III-C produce a selected feature vector $\mathbf{f}_n$ for each node $n \in \mathcal{N}$. In the single-node configurations, $\mathbf{f}_n$ is used directly. In the fusion configuration, the selected feature

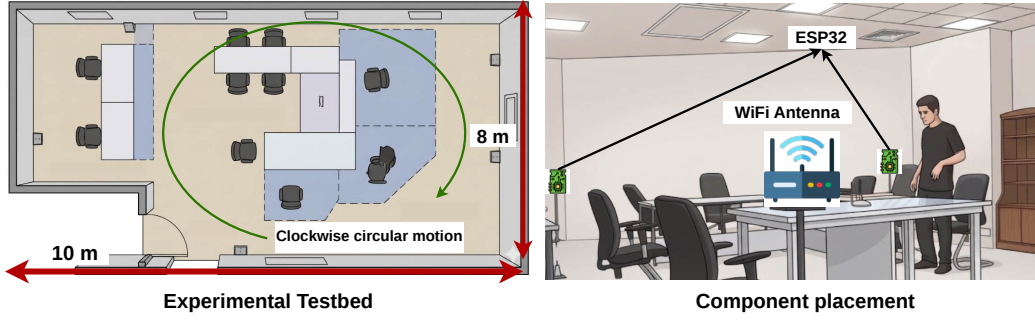


Fig. 1. An illustration of room layout and the positions of the nodes.

vectors of Node 0 and Node 3 are concatenated at the feature level,

$$\mathbf{f}_{\text{fused}} = [\mathbf{f}_0 \parallel \mathbf{f}_3], \quad (9)$$

and the fused vector serves as the classifier input.

E. Classification

Five classifiers are evaluated: Logistic Regression, Support Vector Machine with an RBF kernel, K -Nearest Neighbors, Random Forest, and Gradient Boosting. Each is trained and evaluated under multi node configurations, and their fusion, enabling a direct comparison between single-node sensing and feature-level multi-node fusion.

IV. EXPERIMENTAL EVALUATION

This section evaluates the proposed method. First, the experimental setup describes the dataset, the train and test split, and the metrics (Section IV-A). The evaluation then proceeds in four steps. In the first step, the output of the two-stage feature selection is analysed, reporting how many subcarriers and features were retained and which feature families were selected (Section IV-B). The second step measures the effect of multi-node fusion by comparing the two single-node configurations against their fusion, which tests the main claim of the paper (Section IV-C). The third step repeats this comparison across all five classifiers, to check that the improvement comes from the fusion stage (Section III-D) and not from one specific classifier (Section IV-D). Finally, the per-class behaviour of the best fusion model is analysed using the confusion matrix (Section IV-E).

A. Experimental Setup

The data was collected in a furnished indoor laboratory containing desks, chairs, cabinets, and standard office equipment. The room measured approximately 10×8 m. Figure 1 shows the room layout and the positions of the nodes.

CSI was collected using two low-cost ESP32 nodes as receivers and a commercial Wi-Fi router as the transmitter. The router was placed at the center of the room, and the two ESP32 receivers were positioned at fixed locations around it, mounted on stands at a height of approximately 100 cm, close to the height of the moving subject, and separated by

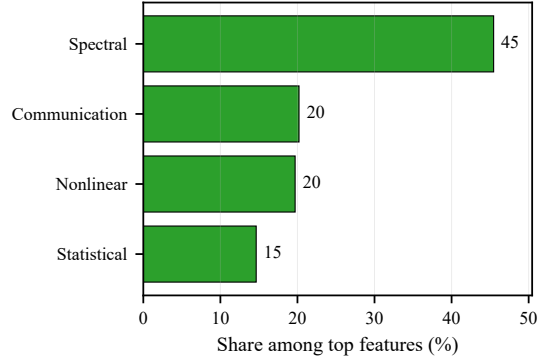


Fig. 2. Share of each feature family among the selected features for the fusion configuration.

approximately 9 m. The link operated on a 20 MHz channel in the 2.4 GHz band, providing CSI over 64 subcarriers per packet at a rate of 170 packets per second. Each activity was recorded for 60 s per trial; because some packets were lost during transmission, the retained dataset is smaller than the nominal recording length.

Three subjects performed five activities (clockwise circular motion, counter-clockwise circular motion, drawing a circle, standing, and sitting), comprising both static postures and dynamic movements. Each continuous recording was divided into non-overlapping segments of $T = 340$ CSI samples, yielding 225 labeled segments per node.

B. Feature Selection Analysis

For the fusion configuration, 55 subcarriers and 320 features were retained. Figure 2 shows how these selected features are distributed across the four feature families. Spectral features account for the largest share (45%), followed by communication and nonlinear features (20% each), and statistical features (15%). The large share of spectral features is expected because different activities produce different frequency patterns in the CSI signal. The fusion configuration also kept more features than either single node, which fits the idea that the two nodes provide different, non-overlapping information.

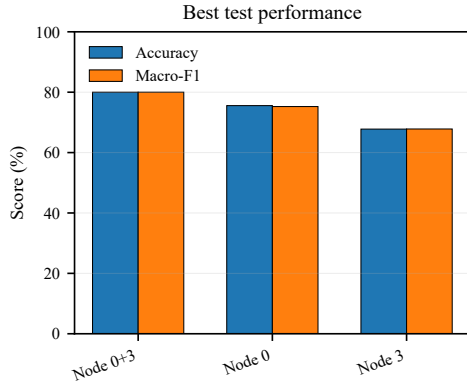


Fig. 3. Accuracy and Macro-F1 of the two single-node configurations and their fusion (best of ten runs).

C. Effect of Multi-Node Fusion

Figure 3 compares the three configurations. The fusion of Node 0 and Node 3 gives the highest performance, with an accuracy of 80.00% and a Macro-F1 of 80.01% using an SVM with an RBF kernel. The best single node, Node 0, reaches 75.56% accuracy and 75.25% Macro-F1, and Node 3 reaches 67.78% and 67.82%. Fusion therefore improves the accuracy by 4.4 percentage points over the best single node.

Reported values are means over ten randomized repetitions; the fusion-vs-best-single-node improvement is confirmed by a paired t-test across the ten runs ($p < 0.05$). Node 3 is clearly weaker than Node 0 when used alone. Fusing the two nodes, however, improves the result over Node 0 rather than lowering it. The weaker node still adds information that the stronger node does not capture, which supports the idea that nodes at different positions observe the same activity differently. These results are obtained on three subjects in a single environment, and broader validation is needed before generalizing the conclusion.

D. Robustness Across Classifiers

Figure 4 shows the average Macro-F1 of each classifier over the ten repetitions. Fusion gives the best result for four of the five classifiers, with the largest improvement for the SVM-RBF classifier (from 64.1% to 71.3%). The exception is Gradient Boosting, for which Node 0 alone is slightly better than fusion. Node 3 is the weakest in all three configurations, and the best result overall is obtained by fusion with the SVM-RBF classifier.

E. Per-Class Analysis

Figure 5 shows the confusion matrix of the best fusion model. Labels 1–5 denote clockwise circular motion, counter-clockwise circular motion, standing and drawing a circle with finger, standing, and sitting, respectively. Activities 1, 2, and 5 are recognized well, each with 89% recall. Most of the error comes from activities 3 and 4 where 44% of activity-4 samples are predicted as activity 3, and 17% of activity-3 samples are predicted as activity 4. These two activities are

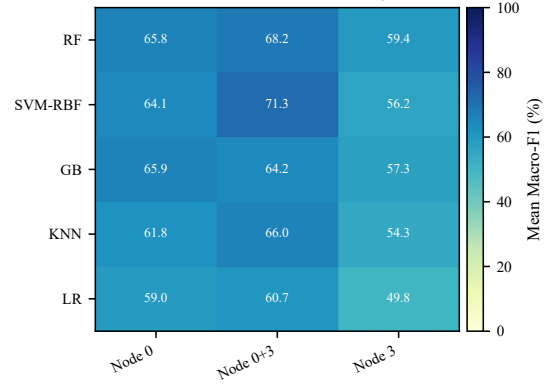


Fig. 4. Average Macro-F1 (%) of each classifier under the three node configurations over ten runs.

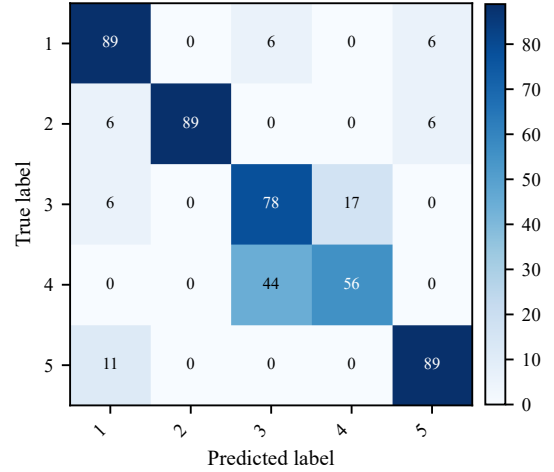


Fig. 5. Confusion matrix of the best fusion model.

similar in movement. The other classes are clearly separated, so the confusion between activities 3 and 4 explains most of the limit on Macro-F1.

F. Comparison with Deep Models

To test whether spatial diversity can substitute for model depth, we compare our interpretable classical-feature pipeline against deep models trained on raw CSI (CNN-1D, ResNet-1D, and MLP) under the same subject-activity-aware protocol, reporting the mean Macro-F1 over ten runs. Figure 6 shows the best classical and best deep model for each node configuration.

On the fused Node 0+3 configuration, the classical SVM-RBF reaches 71.3% Macro-F1, exceeding the best deep model (CNN-1D, 69.1%) by 2.2 percentage points. The classical pipeline also leads on Node 3 (59.4% vs. 56.5%), while the deep model is marginally ahead on Node 0 (68.2% vs. 65.9%). In no configuration does the deep model provide a meaningful advantage, and on the best (fused) configuration the interpretable classifier is strongest overall. This confirms the central claim that spatial diversity from two low-cost nodes substitutes for model depth, removing the data and computational cost of deep models without sacrificing accuracy.

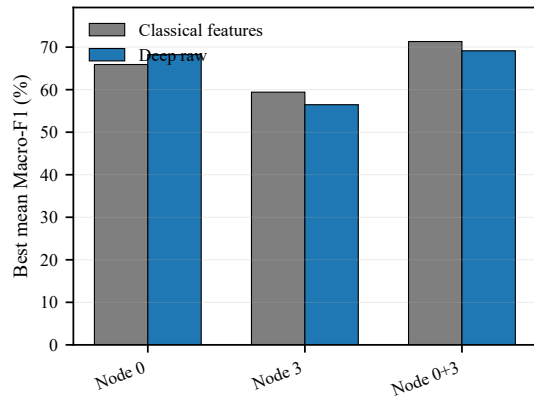


Fig. 6. Best mean Macro-F1 (%) over ten runs: interpretable classical-feature pipeline versus the best deep model trained on raw CSI (best of CNN-1D, ResNet-1D, MLP), per node configuration.

V. CONCLUSION

This paper showed that spatial diversity from multiple low-cost ESP32 nodes can improve WiFi CSI-based Human Activity Recognition without increasing model complexity. Using two spatially separated nodes, a two-stage feature selection scheme, and conventional machine learning, feature-level fusion improved accuracy from 75% for single node to 80%. The weaker node also contributed useful complementary information, indicating that spatially distributed sensing provides a more robust representation than any single node alone.

These results support the use of interpretable multi-node fusion as an alternative to deeper models for addressing the position sensitivity of CSI sensing. The study is limited to two nodes, a controlled environment, and five activities. Future work will investigate larger node deployments, optimized placement strategies, evaluation across more diverse environments, and direct comparison with lightweight deep learning approaches.

REFERENCES

- [1] Akpabio, W., Bulut, E.: Physifi: Wifi sensing for monitoring therapeutic robotic systems. In: 2025 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). pp. 519–524. IEEE Computer Society (2025)
- [2] Armenta-Garcia, J.A., Gonzalez-Navarro, F.F., Caro-Gutierrez, J., Garcia-Reyes, C.I.: Tools and methods for achieving wi-fi sensing in embedded devices. *Sensors* **25**(19), 6220 (2025)
- [3] Choi, H., Fujimoto, M., Matsui, T., Misaki, S., Yasumoto, K.: Wi-cal: Wifi sensing and machine learning based device-free crowd counting and localization. *IEEE Access* **10**, 24395–24410 (2022)
- [4] Gringoli, F., Schulz, M., Link, J., Hollick, M.: 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. pp. 21–28 (2019)
- [5] Haghighat, A., Keshavarz, A., Moradbeikie, A., Lopes, S.I.: Improving rssi fingerprint localization in lorawan networks: An image-based approach. *IEEE Internet of Things Journal* (2026)
- [6] Haghighat, A., Sirous, Z., Moradbeikie, A., Keshavarz, A., Lopes, S.I.: Improving lorawan fingerprint-based localization by detecting and eliminating noisy rssi measurements. In: 2023 IEEE 9th World Forum on Internet of Things (WF-IoT). pp. 1–7. IEEE (2023)

- [7] Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review* **41**(1), 53–53 (2011)
- [8] Hernandez, S.M., Bulut, E.: Lightweight and standalone iot based wifi sensing for active repositioning and mobility. In: 2020 IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks"(WoWMoM). pp. 277–286. IEEE (2020)
- [9] Kang, H., Kim, D., Toh, K.A.: Human activity recognition through augmented wifi csi signals by lightweight attention-gru. *Sensors* **25**(5), 1547 (2025)
- [10] Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials* **15**(3), 1192–1209 (2012)
- [11] Liu, J., Liu, H., Chen, Y., Wang, Y., Wang, C.: Wireless sensing for human activity: A survey. *IEEE Communications Surveys & Tutorials* **22**(3), 1629–1645 (2019)
- [12] Ma, Y., Zhou, G., Wang, S.: Wifi sensing with channel state information: A survey. *ACM Computing Surveys (CSUR)* **52**(3), 1–36 (2019)
- [13] Mekruksavanich, S., Phaphan, W., Hnoohom, N., Jitpattanakul, A.: Attention-based hybrid deep learning network for human activity recognition using wifi channel state information. *Applied Sciences* **13**(15), 8884 (2023)
- [14] Moradbeikie, A., Amorim, I., Azevedo, R., Jesus, C., Beserra, D., Lopes, S.I.: Collaborative indoor positioning: A machine learning approach for dynamic environments with autonomous mobile robots. In: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 6000–6005. IEEE (2025)
- [15] Moradbeikie, A., Azevedo, R., Jesus, C., Beserra, D., Lopes, S.I.: Indoor localization in a collaborative ble-based system through dynamic path loss estimation via location aware autonomous mobile robot. In: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 3542–3547. IEEE (2025)
- [16] Moradbeikie, A., Keshavarz, A., Rostami, H., Paiva, S., Lopes, S.I.: Improvement of rssi-based lorawan localization using edge-ai. In: International Summit Smart City 360°, pp. 140–154. Springer (2021)
- [17] Showmik, I.A., Sanam, T.F., Imtiaz, H.: Human activity recognition from wi-fi csi data using principal component-based wavelet cnn. *Digital Signal Processing* **138**, 104056 (2023)
- [18] Varga, D.: Decision fusion-based deep learning for channel state information channel-aware human action recognition. *Sensors* **25**(4), 1061 (2025)
- [19] Wang, F., Gong, W., Liu, J.: On spatial diversity in wifi-based human activity recognition: A deep learning-based approach. *IEEE Internet of Things Journal* **6**(2), 2035–2047 (2018)
- [20] Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Understanding and modeling of wifi signal based human activity recognition. In: Proceedings of the 21st annual international conference on mobile computing and networking. pp. 65–76 (2015)
- [21] Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., Liu, H.: 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. pp. 617–628 (2014)
- [22] Xie, Y., Li, Z., Li, M.: Precise power delay profiling with commodity wifi. In: Proceedings of the 21st Annual international conference on Mobile Computing and Networking. pp. 53–64 (2015)
- [23] Yang, J., Chen, X., Zou, H., Lu, C.X., Wang, D., Sun, S., Xie, L.: Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing. *Patterns* **4**(3) (2023)
- [24] Yousefi, S., Narui, H., Dayal, S., Ermon, S., Valaee, S.: A survey on behavior recognition using wifi channel state information. *IEEE Communications Magazine* **55**(10), 98–104 (2017)
- [25] Zuo, L.L., Hu, H.M., Qiu, L.Y., Wu, M.W.: In-vehicle csi gesture recognition using two wi-fi receivers with fedtr-lstm model. *IEEE Access* **12**, 113878–113887 (2024)