accepted to appear in Proceedings of 34th International Conference on Computer Communications and Networks (ICCCN), August 4 - 7, 2025, Tokyo, Japan.

MockiFi: CSI Imitation using Context-Aware Conditional Neural Process for Zero-shot Learning

Md Touhiduzzaman and Eyuphan Bulut
Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA
Email: {touhiduzzamm, ebulut}@vcu.edu

***Abstract*—WiFi sensing technology has emerged as a promising technology for activity recognition, leveraging the Channel State Information (CSI) to capture fine-grained movement details. However, the difficulty and scarcity in collecting the required training CSI data with variations hinder the development and deployment of practical WiFi sensing systems in different settings. This paper presents *MockiFi*, a novel system that learns WiFi CSI data transformations across different activities of known individuals and generates CSI data for the activities of new individuals using their base activity CSI data and by mimicking transformations learned from known individuals. Our approach employs a Conditional Neural Process (CNP) to synthesize realistic activity patterns through the learning of pattern transitions using a cosine similarity-based loss function. The effectiveness of our approach is validated through extensive experiments, achieving a high cosine similarity score between the generated and real activity data, indicating the precision and reliability of the generated action fingerprints. We also show that a classifier trained on synthetic data of a new person can successfully recognize their actual activities, demonstrating zero-shot learning capabilities. These results show that MockiFi can help develop customized WiFi sensing systems without the need for collecting excessive new training data, and thus can facilitate their practical usage.**

***Index Terms*—WiFi sensing, synthetic data, data augmentation, human activity recognition, zero-shot learning.**

I. Introduction

The growing capabilities of wireless technologies have paved the way for WiFi sensing as a practical solution for human activity recognition (HAR) [1]. Unlike wearable sensors [2] and vision-based systems [3], WiFi sensing offers a seamless, non-intrusive, and privacy-preserving alternative for HAR through leveraging the Channel State Information (CSI) to detect fine-grained body movements.

WiFi CSI captures detailed changes in wireless signals as they travel from a transmitter to a receiver, reflecting the unique characteristics of the propagation environment. By analyzing subtle variations in wireless signals caused by body movements, CSI enables accurate detection and classification of human activities in WiFi sensing.

While CSI-based systems are promising, they face significant challenges in scalability due to the person- and temporal-specific [4] variations. Collecting diverse training samples for every new user or setting is often costly and impractical. The former refers to the differences in CSI patterns when different individuals perform the same activity, as each person's physical characteristics and movement style can alter the signal in unique ways. On the other hand, temporal effects encompass the variations in CSI data when the same individual performs an activity multiple times, potentially under changing environmental conditions or at different time intervals. These issues hinder the deployment of robust and generalizable WiFi sensing systems.

Fig. 1: Learning CSI transformations across the activities (e.g., walking, running, and crawling) of known users (i.e., Person A and B) can help obtain the CSI fingerprint of activities of a new person (Person C) from a based activity.

To tackle these challenges and develop a generalized solution, several approaches have been studied recently. These approaches leverage principles from zero-shot [5] or few-shot learning [6], [7], transfer learning [8], federated learning [9], generative adversarial networks (GANs) [10] and variational auto encoders (VAE) [11]. However, these systems either use one or a few samples from the unknown activities of people or use some additional information (e.g., text semantics) to relate the unknown activities to known activities (e.g., running is a faster version of walking [12]).

In this paper, we present *MockiFi*, a novel zero-shot learning framework that generates synthetic CSI data for unseen activities of new individuals. By learning transformation patterns between activities from known users, as illustrated in Fig. 1, MockiFi uses a context-aware Conditional Neural Process (CNP) to simulate realistic CSI for unknown cases using just a base activity. Experimental results show that models trained on MockiFi-generated data perform comparably to those trained on real data, enabling scalable and user-adaptive WiFi sensing systems with minimal data collection overhead.

The rest of the paper is organized as follows. We provide a background on WiFi sensing and discuss related works in Section II. In Section III, we present our targeted problem and the details of the proposed MockiFi system. We evaluate the proposed system through experiments in Section IV. Finally, in Section V, we provide some future directions and concluding remarks for this study.

II. Preliminaries

A. WiFi Sensing and CSI

WiFi sensing method utilizes CSI data over the WiFi subcarriers [1], [13]. The CSI matrix is formed with the sum of multiple paths that the signal propagates between a transmitter and a receiver and it is formulated as:

$$H(t) = \sum_{i=1}^{N} \alpha_i(t) e^{-j2\pi f \frac{d_i(t)}{c}}$$

where N represents the number of paths, $d_i(t)$ denotes the length of the i-th path, $\alpha_i(t)$ is the complex variable that consists of the phase and amplitude attenuation information, f is the carrier frequency, and c is the speed of light. The CSI data is a complex number from which we can extract amplitude and phase values, which are then used to train a machine learning model to learn the unique fingerprints of targeted classes. This process can also require some pre-processing steps such as smoothing and anomaly removal, as well as calibration and offset removal in particular for phase values [14].

In this study, we use ESP32 microcontrollers as TX and RX devices to setup our system and use ESP32-CSI-tool [15], [16] to extract the CSI data from the receiver. We then extract and use only the amplitude values over subcarriers to develop the proposed system.

B. Related Work

WiFi sensing has been extensively studied in various applications including but not limited to localization [17], human activity and gesture recognition [18], occupancy monitoring [19], security [20] and health sensing [21]. In this part, we provide an overview of the most related studies to our work in this paper.

CSI Data Augmentation: Augmentation of training data can help increase the accuracy and efficiency of a machine learning model by providing a variety to training data. This is widely applied for image data, but in recent studies [22]–[25], we also see augmentation techniques for CSI data. Most of these techniques are borrowed from computer vision domain thus CSI data is first converted to spectrogram images for proper application of these techniques. Leveraging the principles of zero-shot [5] and few-shot learning [6], synthetic CSI samples are also generated to reduce the volume of the real data needed to train models. This is achieved usually by leveraging additional information such as the word and attribute embeddings from the text domain as it is utilized in [12]. Some other studies have also used GANs [26] or their variants (e.g., conditional GAN [27]) to generate synthetic CSI data. In another approach [28], analysis from online videos of human activities are utilized to teach human activity sensing capabilities without any field measurements.

Conditional Neural Process: Conditional Neural Process (CNP) [29] is a family of neural models structured as neural networks but inspired by the flexibility of stochastic processes, allowing them to make accurate predictions from a handful of training data points while scaling to complex functions and large datasets. In contrast to other generative models such as VAE and GAN, CNP can perform both as a classifier [30] and a generative [31] model with the proper given context. Such context-aware [32], [33] systems are proven to be less resource exhaustive in generating synthetic data for computer vision [31] as well as in natural language processing [33]. Contrastive [34] and evidential [35] variations of CNP show that we can design encoders and decoders to satisfy classification or data augmentation needs. The contextual awareness and the advantage of CNP upon VAE and GAN with WiFi CSI data inspire our work.

Latent Space: Latent space represents a compressed, lower-dimensional embedding of complex data, enabling generative models, like VAEs and GANs, to learn meaningful representations by mapping high-dimensional inputs to a continuous, semantically structured space [36]. While widely applied in computer vision [37] and natural language processing [38] for generating novel content, the potential of latent space exploration remains largely untapped for WiFi CSI data, presenting a promising frontier for signal processing and generative modeling [39].

Contributions: Unlike prior work, this study focuses on learning activity-to-activity transformations from known individuals and applying them to generate unseen activity CSI data for new users using only a base activity. By leveraging a context-aware Conditional Neural Process (CNP), MockiFi enables zero-shot generation of activity fingerprints, offering a novel and practical method for synthetic CSI data generation in WiFi-based human activity recognition.

III. Proposed System

To generate unseen activity CSI for a person, our system uses a known base action and transformation patterns learned from others. Assuming action transitions are consistent across individuals, we formulate the problem and present the implementation.

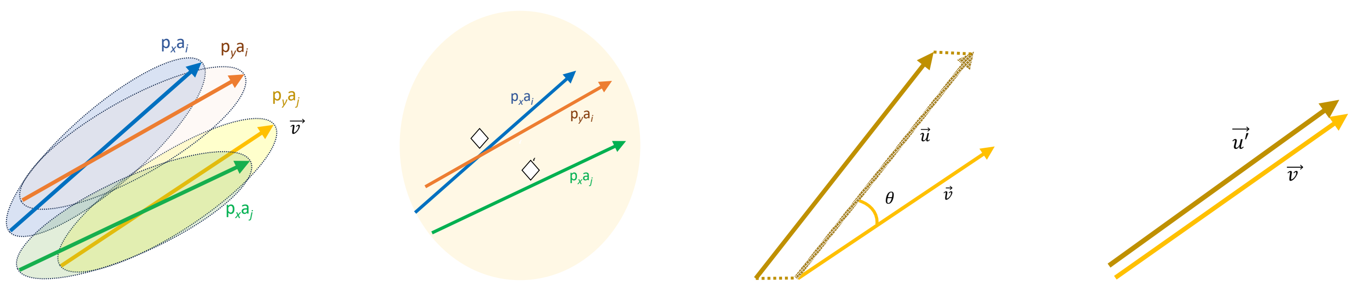
A. Problem Statement

Given a set of persons P and actions A, let $p_x \in P$ perform actions $a_i, a_j \in A$, and $p_y \in P$ perform only a_i. Our goal is to generate $p_y a_j$, the CSI fingerprint of p_y performing a_j, even though $p_y a_j$ is completely unknown — a zero-shot scenario.

We assume that transitions between actions are consistent across individuals. Thus,

$$\begin{aligned} p_y a_j \lozenge p_y a_i \;&\sim\; p_x a_j \lozenge p_x a_i \\ or, p_y a_j \;&\sim\; p_x a_j \lozenge p_x a_i \lozenge' p_y a_i \end{aligned} \tag{1}$$

Leveraging this relationship, a generative model can learn the transformation $\lozenge$ from p_x's activities and apply it to $p_y a_i$ via $\lozenge'$ to synthesize $p_y a_j$.



(a) Latent Space (Hypothetical Plane)

(b) Combining Known Latent Vectors into $\vec{u}$

(c) Match directions of $\vec{u}$ and $\vec{v}$ producing $\vec{u'}$

(d) Match magnitudes of $\vec{u'}$ and $\vec{v}$

Fig. 2: Development of CNP to find the latent vector ($\vec{v}$) by minimizing θ and matching magnitudes using loss functions.

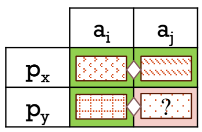


Fig. 3: Targeted contextual relation to learn.

B. Data Augmentation Approach

1) Preprocessing: Following [13], we use only CSI amplitudes, the most common feature in WiFi sensing. To denoise, we apply the Hampel filter and a moving average across non-null subcarriers from our ESP32 device. We then apply PCA to extract major components and use sliding windows of size w over each component, enhancing temporal resolution and increasing the dataset size by a factor of w. These steps are uniformly applied to both generative and classifier models.

2) Normalization in Latent Space Using Loss Functions: We train the Conditional Neural Process (CNP) model, f_θ, using the triplet of actions described in Eq. 1, learning a latent representation $\vec{u}$ aligned with the target $\vec{v}$ of $p_y a_j$. A hypothetical presentation of the latent spaces for the four mentioned CSI sets is shown in Fig. 2a. The overlapping regions represent the confusion of the classifier between the two persons and between the two actions. The tentative latent vectors can be extrapolated as in Fig. 2b. If the angle between these vectors is θ, the cosine of the angle gives us the following relationship:

$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|}.$$

Minimizing θ aligns the directions of $\vec{u}$ with $\vec{v}$, which can be achieved by minimizing the cosine similarity loss defined as:

$$\mathcal{L}_{\text{c}} = 1 - \frac{\sum_{i=1}^{n} \hat{y}_t^i \cdot y_t^i}{\|\hat{y}_t\| \|y_t\|}. \tag{2}$$

Here, $\hat{y}_t^i$ and y_t^i represent the i-th elements of the predicted and ground-truth latent vectors at time t, respectively. While training the data generation model with this cosine similarity loss encourages alignment in the directions of the latent vectors (Fig. 2c), the magnitudes of the vectors also need to be matched (Fig. 2d). To enforce that, we introduce a range consistency loss:

$$\mathcal{L}_{\text{r}} = \sum_{i=1}^{n} \left(\frac{\max(y_{t,:,i}) - \max(\hat{y}_{t,:,i})}{\max(y_{t,:,i})} \right)^2 + \sum_{i=1}^{n} \left(\frac{\min(y_{t,:,i}) - \min(\hat{y}_{t,:,i})}{\max(y_{t,:,i})} \right)^2. \tag{3}$$

The maximum and minimum in this formula are calculated over all rows of each principal component i among n for any time step or CSI window, t. We take the weighted sum with weights λ_c and λ_r for $\mathcal{L}_c$ and $\mathcal{L}_r$ respectively. These weights can be optimized as hyperparameters of the system. Therefore, the final loss function can be defined as

$$\mathcal{L} = \lambda_c \mathcal{L}_c + \lambda_r \mathcal{L}_r. \tag{4}$$

It is noteworthy that this loss function is designed from the vector representations and is differentiable. Therefore, we can optimize our generative model by optimizing this loss function.

3) Context-Aware Conditional Neural Process: The CNP model takes the known actions of one person as context $\mathcal{X}_c$ and a base action of a new person as input x_t, learning the mapping:

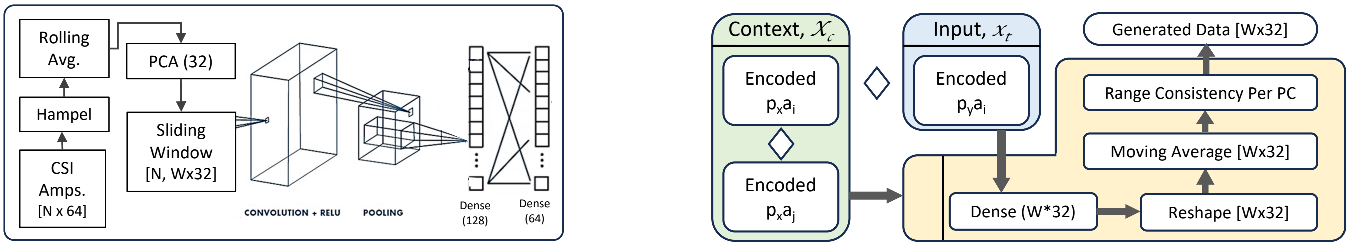
$$f_\theta : (\mathcal{X}_{\text{context}}, x_t) \mapsto \hat{y}_t. \tag{5}$$

This mapping is performed in two steps: (i) encoding and (ii) decoding. For the encoder g_ϕ, we define a one-dimensional convolutional layer of kernel size three and use max-pooling with pool size two after it. Two dense layers of 128 and 64 hidden nodes follow the convolutional layer. The preprocessing steps followed by these neural processes are depicted in Fig. 4a. Hereby, the encoder g_ϕ processes $\mathcal{X}_c$ into latent vector u':

$$u' = g_\phi(\mathcal{X}_{\text{c}}), \tag{6}$$

which, along with encoded x_t, is passed to decoder h_ψ to produce $\hat{y}_t$:

$$\hat{y}_t = h_\psi(u', x_t). \tag{7}$$



(a) Encoder for a single context along with pre-processing.

(b) Decoder merging all context outputs with the encoded input.

Fig. 4: Encoder and decoder networks for the CNP model.

Since the generated CSI data spans short durations (typically under one second), it often contains discrete noise. Although the range consistency loss $\mathcal{L}_r$ helps suppress extreme outliers, we apply an additional denoising step—a moving average over a few iterations per principal component—to smooth the output. Moreover, due to temporal effects during data collection [40], outputs may still exhibit scaling or shifting in the latent space. To correct this, we reapply the range consistency mechanism from $\mathcal{L}_r$ across each principal component and CSI frame. This step is integrated into the decoder, completing the data generation pipeline as illustrated in Fig. 4b.

The CNP model achieves minimal θ^*, ϕ^*, ψ^*, through optimizing the loss function $\mathcal{L}$, the encoder g_ϕ, and the decoder h_ψ, respectively. This optimization process (i.e., $\mathbb{E}$) requires training of CNP by minimizing the total loss over all target data, x_t, as in

$$\theta^*, \phi^*, \psi^* = \underset{\theta,\phi,\psi}{\operatorname{argmin}} \, \mathbb{E}_{\mathcal{X}_c, x_t} [\mathcal{L}] \,. \tag{8}$$

Summary of steps:

1) Preprocess CSI data using Hampel filter, moving average, PCA, and windowing to form $\mathcal{X}_c$ and x_t.
2) Encode $\mathcal{X}_c$ using g_ϕ to obtain u'.
3) Decode u' and encoded x_t via h_ψ to get $\hat{y}_t$.
4) Apply denoising and range normalization to $\hat{y}_t$.
5) Backpropagate and update the parameter sets ϕ and ψ to minimize $\mathcal{L}$.

IV. Evaluation

We evaluate MockiFi on a dataset of five activities performed by seven volunteers. CSI frames were transmitted at 100Hz using low-cost ESP32 devices running the ESP32-CSI-Tool [16] placed at hip level (Fig. 5). Data are collected via a Raspberry Pi 4B. The actions include: walking (a_1), standing run (a_2), raising both arms with joy (a_3), right-hand waving (a_4), and chair sit-stand (a_5).

A. Dataset Splitting and Generation

Each participant performed every activity ten times. We used six repetitions for training and four for testing, applying PCA on training data and transforming test data accordingly. As shown in Fig. 6, the CNP model is trained using all five actions of p_1 and p_2, and generates a_2–a_5 for p_3–p_7 using

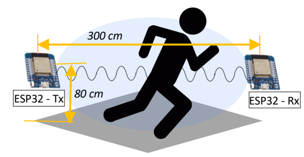


Fig. 5: Experimental setup for CSI data collection.

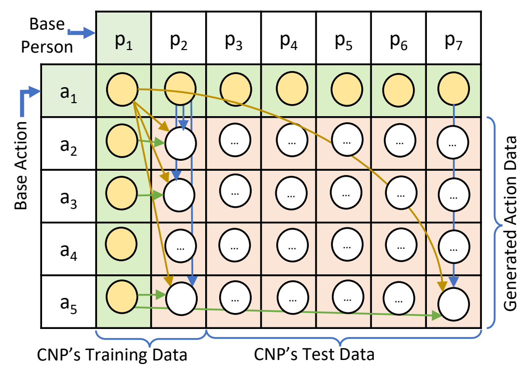


Fig. 6: Data generation flow and train-test split across p_1–p_7 and a_1–a_5.

a_1 as the base action. We evaluate the generated samples by comparing them with the original data of these unseen combinations.

B. Generation Quality via Similarity Metrics

After training the CNP, we generated CSI fingerprints for 20 unseen person-activity combinations. Fig. 7 shows cosine similarity scores between generated and real data, ranging from 0.7576 to 0.9589 (avg. 0.8502), corresponding to an angular deviation (θ) between 16.3° and 40.8° (avg. 31.5°). We also report Jaccard similarity (avg. 0.47), indicating directional alignment and range consistency in the latent space.

C. Classifier-Based Validation

We use a simple CNN (1D convolution + 3 dense layers with 256, 128, and 64 units) for activity and person classification. The intermittent dropout layers have a dropout rate of 0.5. A CSI window of size 150 is used as input, and the Adam optimizer with a learning rate of 0.01 and categorical cross-entropy loss function is employed to train the model over 300 epochs.

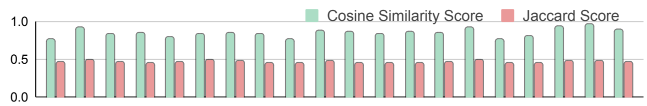


Fig. 7: Cosine (avg. 0.85) and Jaccard similarity (avg. 0.47).

	Base Action				
	a_1	a_2	a_3	a_4	a_5
p_3	72.2	74.5	77.6	76.2	79.3
p_4	73.4	76.3	78.4	78.3	81.2
p_5	71.3	77.4	79.4	81.2	80.2
p_6	68.3	75.3	80.2	81.6	83.2
p_7	70.9	76.3	78.6	80.5	78.8

TABLE I: Activity classification on real data (trained on original data of p_1 and p_2, and the generated data of others).

1) Activity Classifier: We train a single activity classifier for all seven participants. Overall accuracy on the five activities using original data is 78.24% (Fig. 8a), with 81.75% accuracy specifically on the four target activities (a_2–a_5). When the same model is tested on generated samples of those four activities, it yields 78.79% accuracy (Fig. 8b). The 2.96% drop is minor, confirming that generated data closely mimics the statistical properties of real activity fingerprints.

2) Person Classifier: For person identification, the classifier achieves 82.32% accuracy on the original data for all seven individuals. When restricted to p_3–p_7, original data yields 84.11%, and the generated data yields 82.35%, a 1.76% reduction. These results suggest that the generated data retain identity-specific characteristics essential for classification. More complex models or richer datasets could likely improve both metrics further.

Note that while higher accuracy may be achieved with more complex models and more training data, our objective is to replicate unseen activity patterns using learned transformations. In that context, the close alignment between predictions on real and generated data confirms that this goal has been achieved.

D. Zero-shot Learning Performance

To assess MockiFi's generalization, we train classifiers using real CSI from p_1 and p_2 and synthetic data for p_3 to p_7, without exposing real data from these five users during training. Testing is performed on the original data of all participants.

As shown in Table I, activity classification accuracy on real data ranges from 68.3% to 83.2%, depending on person and base action. For example, p_6 scores lowest with a_1 as base (68.3%) and highest with a_5 (83.2%). Accuracy variation per person remains within 5.4%, demonstrating base-invariance.

For person classification (Fig. 9), results range from 72.9% to 82.1%, closely aligning with performance when trained on real data. These results confirm that MockiFi-generated data enables effective zero-shot learning—generalizing to unseen users and actions without requiring labeled samples from them.

V. Conclusion

In this paper, we have explored the possibility of generating realistic CSI fingerprints of new and unknown activities of people using one of their activity data and the CSI transformations learned across different activities from other people. To this end, we proposed *MockiFi*, a context-aware CNP-based framework for zero-shot CSI data generation. MockiFi enables scalable WiFi sensing by synthesizing activity data for new users from a base activity, eliminating the need for exhaustive data collection. Experiments confirm the effectiveness of the generated data in training classifiers, showing promise for adaptable and practical HAR systems. However, looking ahead, future research has several exciting directions, like exploring generative models (e.g., cGAN, CycleGAN, VAE, and so on) and enhancing robustness to environmental variation.

References

[1] F. Miao, Y. Huang, Z. Lu, T. Ohtsuki, G. Gui, and H. Sari, "Wi-Fi Sensing Techniques for Human Activity Recognition: Brief Survey, Potential Challenges, and Research Directions," *ACM Computing Surveys*, 2024.

[2] S. Tan, Y. Ren, J. Yang, and Y. Chen, "Commodity WiFi sensing in ten years: Status, challenges, and opportunities," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 832–17 843, 2022.

[3] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.

[4] C. Chen, G. Zhou, and Y. Lin, "Cross-Domain WiFi Sensing with Channel State Information: A Survey," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–37, 2023.

[5] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, 2019, pp. 313–325.

[6] L. Zhao, R. Xiao, J. Liu, and J. Han, "One is Enough: Enabling One-shot Device-free Gesture Recognition with COTS WiFi," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2024, pp. 1231–1240.

[7] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 540–554, 2020.

[8] Y. Fang, B. Sheng, H. Wang, and F. Xiao, "WiTransfer: A cross-scene transfer activity recognition system using WiFi," in *Proceedings of the ACM turing celebration conference-China*, 2020, pp. 59–63.

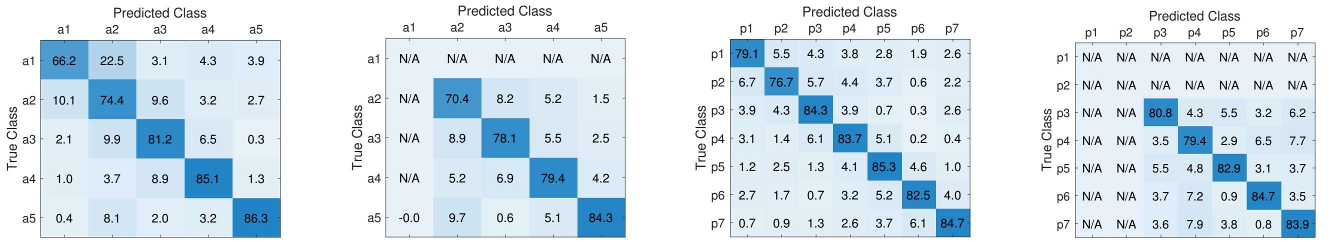
[9] S. M. Hernandez and E. Bulut, "WiFederated: Scalable WiFi Sensing Using Edge-Based Federated Learning," *IEEE Internet Things Journal*, vol. 9, no. 14, pp. 12 628–12 640, 2022.

[10] J. Zhang, Z. Chen, C. Luo, B. Wei, S. S. Kanhere, and J. Li, "MetaGanFi: Cross-domain unseen individual identification using WiFi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–21, 2022.

[11] X. Chen, H. Li, C. Zhou, X. Liu, D. Wu, and G. Dudek, "Fidora: Robust WiFi-based indoor localization via unsupervised domain adaptation," *IEEE IoT Journal*, vol. 9, no. 12, pp. 9872–9888, 2022.

[12] M. T. Islam and S. Nirjon, "Wi-Fringe: Leveraging Text Semantics in WiFi CSI-Based Device-Free Named Gesture Recognition," in *16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2020, pp. 35–42.

[13] S. M. Hernandez and E. Bulut, "WiFi Sensing on the Edge: Signal Processing Techniques and Challenges for Real-World Systems," *IEEE Commun. Surv. Tutorials*, vol. 25, no. 1, pp. 46–76, 2023.



(a) Original Action Classification (Overall accuracy: 78.24%, a_2-a_5 accuracy: 81.75%).

(b) Generated Action Classification (78.79%). N/A shows predictions for non-generated data.

(c) Original Person Classification (Overall accuracy: 82.32%, p_3-p_7 accuracy: 84.11%).

(d) Generated Person Classification (82.35%). N/A shows predictions for non-generated data.

Fig. 8: Confusion matrices for classifiers with real vs. generated data.

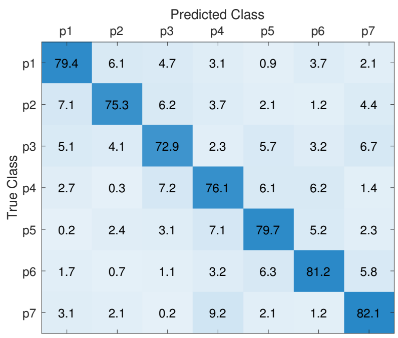


Fig. 9: Person classification (zero-shot) with real test data.

[14] C.-L. Chen, C.-H. Ko, S.-H. Wu, H.-S. Tseng, and R. Y. Chang, "Device-Free Target Following with Deep Spatial and Temporal Structures of CSI," *Journal of Signal Processing Systems*, vol. 95, no. 11, pp. 1327–1340, 2023.

[15] S. M. Hernandez, "ESP32 CSI Tool," 2023. [Online]. Available: https://stevenmhernandez.github.io/ESP32-CSI-Tool/

[16] S. M. Hernandez and E. Bulut, "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)*, Cork, Ireland, Jun. 2020.

[17] J. Xue, J. Zhang, Z. Gao, and W. Xiao, "Enhanced wifi csi fingerprints for device-free localization with deep learning representations," *IEEE Sensors Journal*, vol. 23, no. 3, pp. 2750–2759, 2022.

[18] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *16th Annual Inter. Conference on Mobile Systems, Applications, and Services*, 2018, pp. 401–413.

[19] S. M. Hernandez and E. Bulut, "Adversarial Occupancy Monitoring using One-Sided Through-Wall WiFi Sensing," in *IEEE Inter. Conf. on Communications (ICC), Montreal, Canada, June*, 2021, pp. 1–6.

[20] S. Zhang, R. H. Venkatnarayan, and M. Shahzad, "A WiFi-based Home Security System," in *IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2020, pp. 129–137.

[21] M. Touhiduzzaman, S. M. Hernandez, P. E. Pidcoe, and E. Bulut, "Wi-PT-Hand: Wireless Sensing based Low-cost Physical Rehabilitation Tracking for Hand Movements," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 1, pp. 1–25, 2025.

[22] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-LSTM for human activity recognition using WiFi signal," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4628–4641, 2020.

[23] H. Mo and S. Kim, "A deep learning-based human identification system with wi-fi csi data augmentation," *IEEE Access*, vol. 9, pp. 91 913–91 920, 2021.

[24] W. Hou and C. Wu, "RFBoost: Understanding and Boosting Deep WiFi Sensing via Physical Data Augmentation," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 2, pp. 1–26, 2024.

[25] J. Strohmayer and M. Kampel, "Data augmentation techniques for cross-domain wifi csi-based human activity recognition," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2024, pp. 42–56.

[26] Y. Mao, Z. Guo, B. Sheng, L. Gui, and F. Xiao, "Wi-Cro: WiFi-based Cross Domain Activity Recognition via Modified GAN," *IEEE Transactions on Vehicular Technology*, 2024.

[27] N. N. Bhat, R. Berkvens, and J. Famaey, "CSI4Free: GAN-Augmented mmWave CSI for Improved Pose Classification," in *IEEE 4th International Symposium on Joint Communications & Sensing (JC&S)*, 2024, pp. 1–6.

[28] H. Cai, B. Korany, C. R. Karanam, and Y. Mostofi, "Teaching RF to sense without RF training measurements," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–22, 2020.

[29] M. Garnelo, D. Rosenbaum, C. Maddison, T. Ramalho, D. Saxton, M. Shanahan, Y. W. Teh, D. Rezende, and S. A. Eslami, "Conditional neural processes," in *International conference on machine learning*. PMLR, 2018, pp. 1704–1713.

[30] B. Sheng, F. Xiao, L. Gui, and Z. Guo, "Context-Aware Faster RCNN for CSI-Based Human Action Perception," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 2, 2023.

[31] H. Cai, M. Li, Q. Zhang, M.-Y. Liu, and S. Han, "Condition-aware neural network for controlled image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 7194–7203.

[32] D. Vander Mijnsbrugge, F. Ongenae, and S. Van Hoecke, "Context-aware deep learning with dynamically assembled weight matrices," *Information Fusion*, vol. 100, p. 101908, 2023.

[33] X. He, "Context-aware natural language understanding and generation," Ph.D. dissertation, Monash University, 2022.

[34] Z. Ye and L. Yao, "Contrastive conditional neural processes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9687–9696.

[35] D. S. Pandey and Q. Yu, "Evidential conditional neural processes," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 8, 2023, pp. 9389–9397.

[36] D. P. Kingma, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[37] X. Liu, Y. Zou, L. Kong, Z. Diao, J. Yan, J. Wang, S. Li, P. Jia, and J. You, "Data augmentation via latent space interpolation for image classification," in *Proceedings of the 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 728–733.

[38] Y. Zhou, X. Wang, Y. Niu, Y. Shen, L. Tang, F. Chen, B. He, L. Sun, and L. Wen, "DiffLM: Controllable Synthetic Data Generation via Diffusion Language Models," *arXiv preprint 2411.03250*, 2024.

[39] J. Fonseca and F. Bacao, "Tabular and latent space synthetic data generation: a literature review," *Journal of Big Data*, vol. 10, no. 1, p. 115, 2023.

[40] H. Lee, C. R. Ahn, and N. Choi, "Toward single occupant activity recognition for long-term periods via channel state information," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 2796–2807, 2024.