GenFi: Enhancing WiFi-based Human Activity Recognition to Unseen Scenario via Feature Disentanglement and Meta-Learning

Han Liu[1], Yi Li[1,2], and Biplab Sikdar*[1]

[1]*Department of Electrical and Computer Engineering* , *National University of Singapore*, Singapore
[2]*School of Automation* , *Central South University*, China
Email: {han.liu@u.nus.edu, yli_csu@csu.edu.cn, bsikdar@nus.edu.sg}

***Abstract*—WiFi-based sensing technology has gained significant attention for its ability to enable pervasive human activity recognition (HAR) in indoor spaces. One challenge is that current WiFi HAR systems often experience performance degradation in unseen scenarios (e.g., new environments, people, and weather conditions). Some research has attempted to address this issue by extracting scenario-invariant features using deep learning (DL). However, discarding scenario-specific features brings about insufficient representation of task-related information, resulting in limited model adaptability. In this paper, we present GenFi, a robust WiFi HAR system that enhances model generalization by leveraging both scenario-invariant and scenario-specific features. To achieve this, GenFi first disentangles the raw input into these two types of features through adversarial learning and correlation analysis. Subsequently, GenFi uses meta-learning to self-optimize the fusion of these two features, leading to a generalized cross-scenario WiFi HAR system. Compared to state-of-the-art approaches, GenFi achieves the best trade-off between high performance and low complexity in diverse unseen scenarios, making it a promising solution for real-world deployment.**

***Index Terms*—WiFi Sensing, Human Activity Recognition, Domain Generalization, Feature Disentanglement, Meta-learning.**

I. Introduction

With the rising popularity of WiFi technology and the broad availability of WiFi infrastructure, WiFi sensing is drawing increasing attention in emerging fields such as Internet of Things (IoT) systems and healthcare services [1], [2]. Compared to traditional vision-based sensing technologies, WiFi sensing is device-free, low-maintenance, and privacy-preserving. Even in challenging conditions such as dark environments and non-line-of-sight (NLOS) situations, WiFi-based sensing systems still operate effectively. These advantages enable WiFi sensing technology to be a promising solution for human activity recognition (HAR) in homes, offices, and other indoor spaces [3], [4]. As deep learning (DL) shows significant success in recognition tasks, DL-based WiFi HAR has attracted growing attention [5]. A major challenge is that DL-based HAR systems often face performance degradation when applied to new scenarios, as they depend heavily on training data [6]. Considering that real-world application scenarios are dynamic and diverse, solving this problem is crucial for the practical deployment of DL-based WiFi HAR systems.

Recently, some methods have been proposed to adapt DL-based WiFi HAR systems to unseen scenarios, using channel state information (CSI) as training data [7]–[11]. Given that CSI contains multi-layered semantic information such as weather, environment, person, and object motion [12], the core idea of these methods is to extract scenario-invariant human-motion features, which helps the model focus on generalizable patterns rather than overfitting to scenario-specific information [13], [14]. For example, [7]–[9] employ domain adaptation and domain generalization techniques to learn scenario-invariant features. In [7], [8], the pre-trained models need to be tailed with at least one sample from new scenarios, which increases the difficulty of their practical deployments. AirFi [9] extracts scenario-invariant features by minimizing the distribution differences among CSI from different previously seen environments, without requiring new data. Nevertheless, It focuses only on environmental dynamics and neglects factors such as people and weather conditions, leading to limited applicability. Approaches based on Doppler shift analysis [10], [11] demonstrate strong robustness to diverse variations, with no new data required. However, Doppler analysis involves complex signal correction and iterative optimizations, making it inefficient for real-time inference.

Furthermore, relying solely on scenario-invariant features may be insufficient for scenario adaptation, as it may lead to severe information loss, especially when there is a large distribution gap between the source and target scenarios [15]. Properly incorporating scenario-specific information can be beneficial for cross-scenario recognition, which has been explored in the field of computer vision [16], [17]. Therefore, we believe that WiFi-based HAR models can achieve better generalization by leveraging scenario-specific information. The challenge lies in balancing reliance on scenario-specific information: too little results in insufficient representation, while too much leads to overfitting. To address these problems, we propose GenFi, a novel WiFi-based HAR system that can be easily generalized to diverse unseen scenarios, by considering both

This research is supported by A*STAR, CISCO Systems (USA) Pte. Ltd and National University of Singapore under its Cisco-NUS Accelerated Digital Economy Corporate Laboratory (Award I21001E0002).

scenario-invariant and scenario-specific information. Specifically, GenFi first uses feature disentanglement techniques to isolate the scenario-invariant and scenario-specific features from raw CSI by encouraging the independence between them [18]. Following this, GenFi adopts meta-learning [19] to self-learn how to fuse scenario-specific and scenario-invariant features for robust cross-scenario performance.

The contribution of this paper is as follows:

- We propose GenFi, a WiFi-based HAR system that can be efficiently generalized to diverse unseen scenarios including new environments, people, and weather conditions, with no new data required. Compared to previous approaches, GenFi has greater applicability, robust performance, and lower implementation complexity.
- We consider both scenario-invariant and scenario-specific features to enhance model generalization. To the best of our knowledge, GenFi is the first work that fuses scenario-specific features with scenario-invariant features to achieve a robust cross-scenario WiFi HAR system.
- We explore an efficient training strategy for feature fusion. By applying feature disentanglement, scenario-invariant and scenario-specific features are separated from raw CSI. Leveraging meta-learning strategy, GenFi self-learns how to fuse these features effectively to enhance the generalization capability of the system.
- We employ efficient preprocessing and augmentation techniques to improve data representation. Static trends and hardware errors in CSI are eliminated using first-order differencing. Random erasing and channel shuffling are applied to enhance data diversity.

II. Technical Background

In this section, we are going to review feature disentanglement and meta-learning, which are two typical techniques for model generalization.

A. Feature Disentanglement for Model Generalization

Data often contains interwoven features, some of which are stable across scenarios while others are sensitive to scenario changes. Feature disentanglement refers to the process of isolating these distinct types of features, allowing us to extract scenario-invariant features, thereby enhancing model generalization. In recent years, Feature disentanglement has shown significance in improving model robustness across various fields. For instance, in the field of voice conversion, [20] separates content-related and speaker-related features using a pair of encoders, enabling the model to convert speech from unseen speakers into a target voice. In the context of face recognition, [21] introduces a deep adversarial disentangled network to isolate scenario-specific features from identity features, ensuring successful classification of face images captured in different scenarios. In this paper, we use a dual-encoder structure to learn scenario-invariant and scenario-specific features, and disentangle them by minimizing their overlap through adversarial training and correlation analysis.

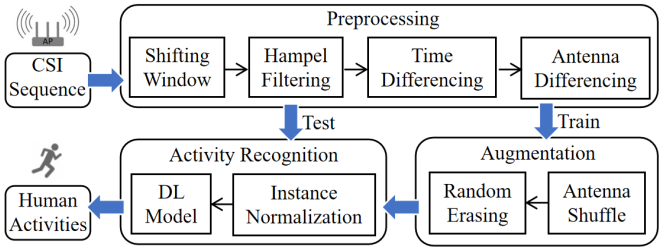

Fig. 1: System overview.

B. Meta-Learning for Model Generalization

Meta-learning aims to provide models with the ability to learn how to learn. It consists of two phases: meta-training and meta-testing. During both phases, the model upgrades its parameters as it learns. In meta-training phase, the model is exposed to diverse data from various domains, allowing it to identify common patterns and acquire generalizable knowledge. During meta-testing, the model's effectiveness is evaluated on unseen data to simulate how it would perform when faced with new scenarios. The testing loss will guide the model's further updates, ultimately facilitating effective model generalization [22]. In this work, leveraging meta-learning strategy, GenFi self-learns how to reasonably integrate scenario-specific information with scenario-invariant features, improving its robustness and reliability in unseen scenarios.

III. System Design

The overall architecture of GenFi is presented in Fig. 1. In this section, we discuss each block of the system in detail, including data preprocessing, augmentation, and model training. During the training process, basic-learning stage (Fig. 2) is conducted first, followed by meta-learning stage (Fig. 3).

A. Data Preprocessing and Augmentation

In practical applications, the receiver continuously records CSI. To create inference samples, we first utilize moving windows to extract CSI slices. Given that the raw CSI data often contains significant noise, Hampel filter is applied to remove outliers. Following that, first-order differencing is performed along time dimension to eliminate static trends and along antenna dimension to alleviate inherent hardware errors.

To improve model robustness, we apply data augmentation to increase the diversity of training samples. Considering the recognition results are influenced by antenna ordering, which depends on the relative positions of the device and person within the room, we randomly shuffle the antenna dimension sequence to mitigate this effect. Additionally, We employ random erasing by masking a portion of the input, promoting a sparser feature representation to avoid overfitting on redundant information. The processed CSI from multiple scenarios will be used for model training. When feeding each sample into the DL model, instance normalization is employed to help the model focus on individual samples from different scenarios, mitigating the impact of distributional differences across scenarios and improving generalization.

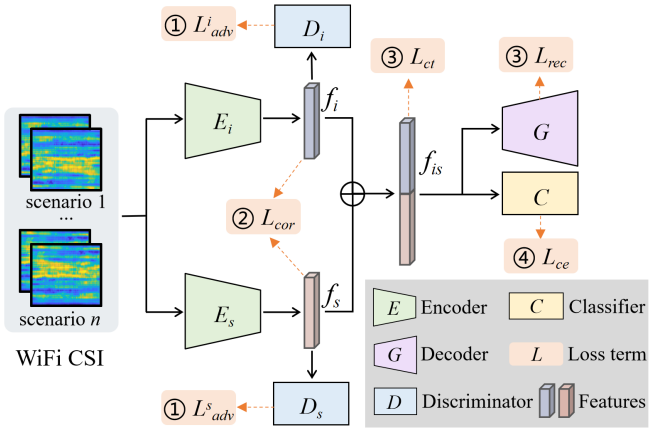

Fig. 2: Basic-learning stage of model training. The subscript i and s denote scenario-invariant and scenario-specific, respectively. Modules with the same color have the same structure.

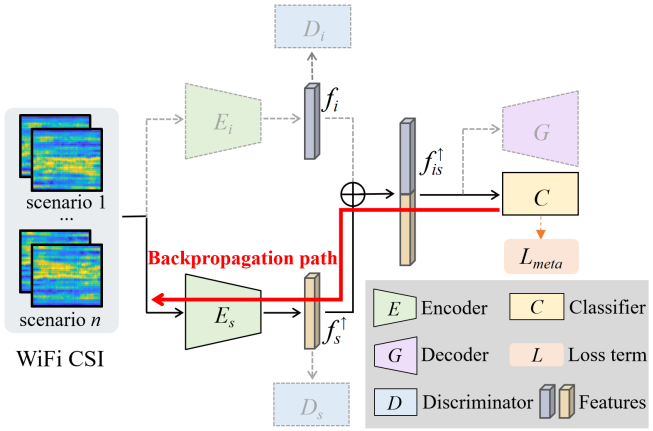

Fig. 3: Meta-learning stage of model training. In this stage, only E_s and C will be updated. Other modules are frozen.

B. Basic-Learning Stage of Model Training

Fig. 2 illustrates the basic-learning stage of model training. The model comprises a scenario-invariant encoder E_i, a scenario-specific encoder E_s, a scenario-invariant discriminator D_i, a scenario-specific discriminator D_s, a decoder G, and a classifier C. We will describe the calculation of loss terms at each step in basic-learning stage in the following part.

1) Feature Extraction via Adversarial Learning: A dual-encoder structure is used to extract scenario-invariant features $f_i = E_i(x)$ and scenario-specific features $f_s = E_s(x)$ in parallel, where x denotes input sample. Each encoder consists of three 2D convolutional layers. To ensure that these two feature codes contain their aimed information with minimal redundancy, we apply an adversarial training framework that includes D_i and D_s, each constructed of two fully connected layers. Specifically, the goal of D_i is to correctly identify the scenario label from f_i, while E_i tries to project the input sample into the latent space in a way that makes it hard for D_i to tell which scenario the input belongs to. For f_s, a symmetrical structure is designed, where both E_s and D_s aim to maximize the prediction probability of scenario label from f_s. The adversarial loss terms $\mathcal{L}^i_{\text{adv}}$ and $\mathcal{L}^s_{\text{adv}}$ are given as below, where z denotes scenario label:

$$\mathcal{L}^i_{\text{adv}} = \min_{E_i} \max_{D_i} -\mathbb{E}_{x,z}[z \log D_i(f_i)], \tag{1}$$

$$\mathcal{L}^s_{\text{adv}} = \min_{E_s, D_s} -\mathbb{E}_{x,z}[z \log D_s(f_s)]. \tag{2}$$

2) Feature Disentanglement by Correlation Analysis: To reduce the overlap between f_i and f_s and encourage their independence, we disentangle them by forcing their correlation matrix to approach 0. Correlation matrix is a square matrix describing the linear relationships between multi-dimensional variables, of which each element is a Pearson correlation coefficient calculated by:

$$\text{Cor}(a,b) = \frac{\text{Cov}(a,b)}{\sigma_a \sigma_b} = \frac{\sum(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum(a_i - \bar{a})^2 \sum(b_i - \bar{b})^2}}, \tag{3}$$

where $\text{Cov}(\cdot)$ and σ represent covariance and standard deviation, respectively. The optimization problem can be solved by minimizing the correlation loss $\mathcal{L}_{\text{cor}}$, given by:

$$\mathcal{L}_{\text{cor}} = \min_{E_i, E_s} \mathbb{E}_x[\|\text{Cor}(f_i, f_s)\|_2^2]. \tag{4}$$

3) Fused Feature Representation Optimization: As discussed earlier, model generalization can be boosted by incorporating scenario-specific information. Therefore, f_i and f_s are concatenated to obtain the fused features f_{is}. The reconstruction loss $\mathcal{L}_{\text{rec}}$ is used for unifying the content of feature codes encoded from different training scenarios, calculated as:

$$\mathcal{L}_{\text{rec}} = \min_{E_i, E_s, G} \mathbb{E}_x[\|G(f_{is}) - x\|_2^2]. \tag{5}$$

Moreover, to emphasize the feature representation associated with different activity labels, the contrastive loss $\mathcal{L}_{\text{ct}}$, which encourages same-class samples to be closer and different-class samples to be farther apart, is employed as follows:

$$\mathcal{L}_{\text{ct}} = \min_{E_i, E_s} \mathbb{E}_{x,y}\left[\|f^y_{is}, f^{y+}_{is}\|_2 + \max(0, m - \|f^y_{is}, f^{y-}_{is}\|_2)\right]. \tag{6}$$

where y and m denote activity label and margin constraint, respectively. $\|f^y_{is}, f^{y+}_{is}\|_2$ represents the Euclidean distance between feature pairs of the same activity label; $\|f^y_{is}, f^{y-}_{is}\|_2$ denotes that between feature pairs of different activity labels.

4) Classifier Optimization: Finally, the optimized fused features f_{is} are fed into the classifier, consisting of a single fully connected layer. The cross-entropy loss $\mathcal{L}_{\text{ce}}$ is used to measure the activity prediction errors as follows:

$$\mathcal{L}_{\text{ce}} = \min_{E_i, E_s, C} -\mathbb{E}_{x,y}[y \log C(f_i, f_s)]. \tag{7}$$

Overall, the loss function $\mathcal{L}_{\text{basic}}$ for the basic-learning stage of GenFi is identified as:

$$\mathcal{L}_{\text{basic}} = \mathcal{L}^i_{\text{adv}} + \mathcal{L}^s_{\text{adv}} + \lambda_1 \mathcal{L}_{\text{cor}} + \lambda_2 \mathcal{L}_{\text{rec}} + \lambda_3 \mathcal{L}_{\text{ct}} + \mathcal{L}_{\text{ce}}, \tag{8}$$

where λ_i denotes the weight parameter for the loss term.

C. Meta-Learning Stage of Model Training

To enable model to learn which scenario-specific features should be focused on, we further adopt meta-learning strategy after basic-learning stage to enhance model's adaptability to new scenarios. During this stage, only E_s and C are updated and the other modules are frozen, as shown in Fig 3. To implement meta-learning, we divide all training data by sequentially selecting data from one scenario as meta-testing subset S_{mte}, with data from the remaining scenarios forming meta-training subset S_{mtr}.

S_{mtr} is utilized for meta-training. E_i is still employed to obtain scenario-invariant features f_i, consistently using the parameters from the last basic optimization. Meanwhile, the scenario-specific features are extracted by the continuously updated E_s. For clarity, we use $f_s^{\uparrow}$ to denote the scenario-specific features within the meta-learning stage. Ultimately, f_i and $f_s^{\uparrow}$ are fused to form the final feature representation and used for updating C. During meta-testing, S_{mte} is utilized to evaluate the model's robustness on an unseen scenario. E_s and C will be further optimized, based on cross-entropy loss in meta-testing phase. The joint update formula for E_s and C in meta-learning stage is:

$$\mathcal{L}_{\text{meta}} = \min_{E_s,C} f(\theta^{E_s,C} - \nabla f(\theta^{E_s,C}, S_{mtr}), S_{mte}), \quad (9)$$

where $\theta^{E_s,C}$ denotes their joint parameters and

$$f(\theta, S) = -\mathbb{E}_{x,y}[y \log C(f_i, f_s^{\uparrow})]. \quad (10)$$

For a batch of training data, the model first performs basic-learning stage according to Eq. (8) across all modules, followed by the meta-learning stage that further updates the scenario-specific encoder and the classifier by Eq. (9).

IV. Evaluation

In this section, we evaluate GenFi in diverse testing scenarios, including performance comparisons, complexity & latency analysis, and ablation study.

A. Experimental Settings

1) Dataset: We evaluate our proposed method with an open dataset [23], which contains CSI-based HAR samples for eight activity classes, involving thirteen participants across seven environments over various days. The dataset is collected over 242 WiFi OFDM sub-channels and includes measurements from four antenna pairs, resulting in the original CSI sequence dimension of $4\times242\times N$, where N denotes the number of sampling points. Each CSI sequence corresponds to a volunteer repeatedly performing a specific action over a period of time. Diverse variations such as environments (e.g., room layout and device position), people (e.g., gender and age), weather conditions (e.g., temperature and humidity), and transmission paths (e.g., line-of-sight (LOS) and NLOS) are considered, enabling simulation of diverse unseen conditions that may arise in real-world WiFi sensing deployments. As we are striving for model generalization, the diversity of the training and testing subsets is taken into account. Accordingly, seven subsets are selected from the dataset, as shown in Table I. In experiments, through various pairs of training and testing set combinations, we comprehensively evaluate the model's performance under different unseen factors.

TABLE I: Characteristics of subsets used for experiments.

Subset	Environment	Person	Time	Path
A	Bedroom	P1	a	LOS
B	Bedroom	P1	b	LOS
C	Bedroom	P2	c	LOS
D	Bedroom	P1	d	NLOS
E	Bedroom	P2	e	NLOS
F	Living room	P1	f	LOS
G	Laboratory	P3	g	LOS

2) Implementation Details: We conduct experiments using PyTorch with one NVIDIA A100 80GB GPU. The size of each CSI frame is $4\times242\times340$, with a shifting window length set to 340 (around two seconds, corresponding to typical period of human motion). After two first-order differencing operations, the shape of a CSI sample is $3\times242\times339$. The loss-related hyperparameters are set as $\lambda_1 = 2$, $\lambda_2 = \lambda_3 = 0.5$, and $m = 1$, which aims to balance the magnitudes of multiple losses. Adam optimizer is used. For GenFi, the learning rate for basic-learning, meta-training, and meta-testing is 10^{-4}, 10^{-4}, and 10^{-5}, respectively. For all baselines, the learning rate is 10^{-4}. For performance evaluation, accuracy is used as the metric. For complexity analysis, we count floating point operations (FLOPs) and record latency which is time for completing one inference including data processing and model prediction.

B. Experimental Results and Analysis

1) Performance Comparisons: We compare the proposed GenFi with recent related work [7], [9], [10], all of which aim to enhance the generalization capability of WiFi sensing systems. The training subsets are fixed as A, D, and G, as they involve different environments (bedroom and laboratory), people (P1 and P3), time periods (a, d, and g), and transmission paths (LOS and NLOS). To test the ability of each method to adapt to different types of unseen scenarios, we take B, C, and F as testing subsets in turn, simulating situations in which the model faces unseen time (b), unseen person (P2), and unseen environment (living room), respectively. The experiment results are shown in Table II, where we highlight the best results (with *) and the second-best results (with bold).

According to Table II, the average prediction accuracy of GenFi exceeds 90% in all testing scenarios. In terms of the accuracy of individual activities, GenFi consistently achieves the best or second-best results in most classes. Such results show the strong reliability and adaptability of GenFi across diverse unseen situations. SHARP achieves excellent performance by estimating human movement velocity from Doppler shift information, which is independent of scenario variation. Even so, we will demonstrate later in IV-B2 that the computing complexity and inference latency of SHARP are significantly higher than those of GenFi, making it inefficient for online

TABLE II: Classification accuracy (%) comparisons of all models under diverse unseen scenarios.

Model	Empty	Sitting	Standing	Sit down stand up	Walking	Running	Jumping	Arm gym	Macro average	Weighted average	Testing scenario
DAFi [7]	58.55	99.71	82.38	91.89	78.39	**88.89**	94.36	76.92	83.88	83.79	
AirFi [9]	30.57	92.02	77.72	87.03	51.26	85.35	83.08	74.36	72.67	72.51	ADG-B
SHARP [10]	100*	**100***	94.56*	**100***	100*	99.84*	**100***	100*	99.30*	99.30*	(unseen time)
GenFi	**84.97**	**100***	**90.67**	**100***	**84.42**	87.88	**100***	**84.10**	**91.51**	**91.40**	
DAFi [7]	77.00	**93.17**	82.05	40.22	85.43	75.62	98.43	89.80*	80.21	80.69	
AirFi [9]	47.50	88.29	86.67	33.52	58.79	59.20	97.95	76.02	68.49	68.85	ADG-C
SHARP [10]	96.07*	100*	93.91*	98.56*	100*	**84.43**	98.72*	**87.58**	94.91*	94.79*	(unseen person)
GenFi	**86.00**	88.78	**91.79**	**93.30**	**98.49**	84.58*	**98.46**	84.18	**90.90**	**90.64**	
DAFi [7]	96.43	98.48	83.16	80.65	69.34	**83.00**	39.29	73.60	78.00	77.95	
AirFi [9]	72.45	93.72	56.12	86.56	67.88	75.50	60.20	72.59	73.13	72.99	ADG-F
SHARP [10]	100*	100*	**90.32**	99.82*	100*	73.16	100*	100*	95.40*	95.32*	(unseen environment)
GenFi	**98.98**	**98.95**	95.92*	**99.46**	**82.90**	85.00*	**90.82**	**79.19**	**91.40**	**91.32**	

real-time applications. DAFi achieves an average accuracy of around 80% in each round of testing, with approximately 10% lower than that of GenFi. This is because, in the original implementation, DAFi requires samples from the target scenario to effectively reduce the distribution gap between source and target scenarios. However, such data are not provided in our experiments to simulate real-world deployment; instead, we replace that with data from the training set, which prevents DAFi from attaining satisfactory performance. AirFi shows poor generalization performance in our experiments, as it focuses solely on environmental variables while ignoring other factors. When more different variable factors are involved, the effectiveness of the domain alignment method used in AirFi is diminished. As shown in Table II, AirFi achieves its best performance in the *unseen environment* testing scenario, with performance degradation in other unseen scenarios, which demonstrates its limitations in adapting to diverse scenarios.

TABLE III: Classification accuracy (%) comparisons of all models under the most challenging scenario.

Model	Macro average	Weighted average	Testing scenario
DAFi [7]	70.72	70.48	
AirFi [9]	67.53	67.33	CEG-F
SHARP [10]	88.70*	88.62*	(all)
GenFi	**83.81**	**83.73**	

In real-world applications, a more common situation is that the model will operate in a new environment, for a new person, at a future time, i.e., all factors are entirely unseen to the model. To test models under this most challenging scenario, we further select C, E, and G as training subsets, which are across different environments (bedroom and laboratory), people (P2 and P3), types of path (LOS and NLOS), and time periods (c, e, and g), and take F as testing subset (living room, P1, f) where all factors are unseen to models. The testing results are shown in Table III. It can be observed that GenFi achieves above 83% classification accuracy on testing data, with about 13% and 16% performance gains over DAFi and AirFi, respectively. SHARP still shows the best performance, achieving 5% improvement over GenFi. Despite this, the computational cost of SHARP is extremely high (see IV-B2), resulting in a worse trade-off between performance and computational efficiency compared to GenFi.

2) Complexity and Latency Analysis: In practical applications, WiFi sensing models are typically deployed on mobile devices, which requires the complexity to be kept at a reasonable level. Meanwhile, low processing latency is crucial for ensuring real-time responsiveness. Based on these considerations, we compared the complexity and latency of different models, as illustrated in Fig. 4, where the units of complexity and latency are FLOPs and seconds, respectively. To simulate the operation of models on lightweight mobile devices, results in this part are obtained based on CPU implementations. Fig. 4 shows that the computing complexity of SHARP is two orders of magnitude, nearly 255 times, higher than that of GenFi, incurring significantly greater computational costs. This is because SHARP involves extensive iterative optimization and regression analysis to estimate Doppler information. In terms of latency, SHARP spends over 120 seconds for a single inference, almost 65 times slower than GenFi, which makes it hard to deploy in real-world applications. In contrast, GenFi can not only significantly outperform DAFi and AirFi, but also maintain comparable complexity and inference latency with them. This indicates that GenFi is well-suited for deployment on lightweight devices for real-time inference.

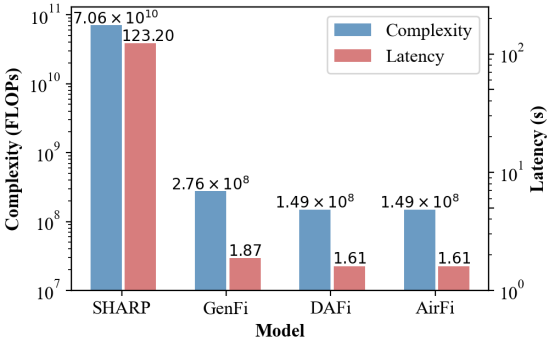

Fig. 4: Complexity and latency comparisons.

TABLE IV: Ablation study on accuracy(%).

Model	Macro average	Weighted average	Testing scenario
GenFi w/o aug	81.35	81.11	
GenFi w/o fusion	83.28	83.05	ADG-B
GenFi w/o meta	90.08	90.04	(unseen time)
GenFi	91.51*	91.40*	
GenFi w/o aug	78.11	78.60	
GenFi w/o fusion	89.25	89.30	ADG-C
GenFi w/o meta	84.87	84.84	(unseen person)
GenFi	90.90*	90.64*	
GenFi w/o aug	84.21	84.12	
GenFi w/o fusion	87.67	87.65	ADG-F
GenFi w/o meta	90.59	90.55	(unseen environment)
GenFi	91.40*	91.32*	
GenFi w/o aug	76.81	76.66	
GenFi w/o fusion	82.33	82.25	CEG-F
GenFi w/o meta	79.05	78.91	(all)
GenFi	83.81*	83.73*	

3) Ablation Study: To study how each module contributes to the generalization capability of GenFi, we further conduct the ablation study, as shown in Table IV, where we highlight the best results (with *). In this table, *GenFi w/o aug* represents GenFi without data augmentation; *GenFi w/o fusion* denotes that we only feed the scenario-invariant feature into the classifier without feature fusion; *GenFi w/o meta* indicates that we directly use concatenated features as the input of the classifier, without the optimization from meta-learning. It can be observed that the performance of GenFi declines when any single module is absent, highlighting the effectiveness of all these techniques employed. Specifically, the missing data augmentation, which includes erasing (to simulate packet loss) and antenna shuffling (to simulate antenna ordering effects), results in the greatest performance degradation. When feature fusion and meta-learning are not utilized, the model's performance becomes unstable across various scenarios, leading to performance drops of up to 8% and 6% in different testing scenarios, respectively. These findings highlight the importance of all modules employed in GenFi for stably adapting the model to diverse scenarios.

CONCLUSION

In this paper, we have presented GenFi, a generalized and easily adaptive channel state information (CSI)-based human activity recognition (HAR) system. By utilizing adversarial learning and correlation analysis, GenFi disentangles the scenario-invariant and scenario-specific features from the raw CSI data. Leveraging meta-learning, GenFi self-learns how to effectively fuse these two types of features to enhance its robustness in diverse unseen scenarios. Our experiments demonstrate that GenFi can achieve excellent performance with low complexity and latency. Compared to existing methods, GenFi achieves up to 16% performance gain in the most challenging testing scenario. Moreover, our analysis reveals that GenFi reduces complexity by 99.6% and latency by 98.5% compared to the best-performing system, with only a 5% performance gap. In conclusion, GenFi effectively balances strong generalization performance with lightweight computational demands and low inference latency, making it highly efficient for real-world deployment.

REFERENCES

[1] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1-36, 2019.

[2] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf WiFi devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2466–2479, 2016.

[3] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, "Wichase: A WiFi based human activity recognition system for sensorless environments," in *Proc. IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 1-6, 2017.

[4] L. Guo, L. Wang, J. Liu, W. Zhou, and B. Lu, "HuAc: Human activity recognition using crowdsourced WiFi signals and skeleton data," *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, p. 6163475, 2018.

[5] I. Nirmal, A. Khamis, M. Hassan, W. Hu, and X. Zhu, "Deep learning for radio-based human sensing: Recent advances and future directions," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp.995-1019, 2021.

[6] 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.

[7] H. Li, X. Chen, J. Wang, D. Wu, and X. Liu, "DAFI: WiFi-based device-free indoor localization via domain adaptation," in *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 4, pp .1-21, 2021.

[8] B. Sheng, R. Han, H. Cai, F. Xiao, L. Gui, and Z. Guo, "CDFi: Cross-Domain Action Recognition using WiFi Signals," *IEEE Transactions on Mobile Computing*, 2024.

[9] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "AirFi: Empowering wifi-based passive human gesture recognition to unseen environment via domain generalization," *IEEE Transactions on Mobile Computing*, vol. 23, no. 2, pp.1156-1168, 2022.

[10] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHAPR: Environment and person independent activity recognition with commodity IEEE 802.11 access points," *IEEE Transactions on Mobile Computing*, vol. 22, no. 10, pp. 6160-6175, 2022.

[11] N. Lyons, A. Santra, V. K. Ramanna, K. Uln, R. Taori, and A. Pandey, "WIFIACT: Enhancing human sensing through environment robust preprocessing and Bayesian self-supervised learning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 13391-13395, 2024.

[12] W. Zhang, S. Zhou, L. Yang, L. Ou, and Z. Xiao, "WiFiMap+: High-level indoor semantic inference with WiFi human activity and environment," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7890-7903, 2019.

[13] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052-8072, 2022.

[14] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4396-4415, 2022.

[15] H. Zhao, R. T. Des Combes, K. Zhang, and G. Gordon, "On learning invariant representations for domain adaptation," in *Proc. International Conference on Machine Learning (ICML)*, pp. 7523-7532, 2019.

[16] W. G. Chang, T. You, S. Seo, S. Kwak, and B. Han, "Domain-specific batch normalization for unsupervised domain adaptation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7354-7362, 2019.

[17] S. Yang, Y. Wang, J. Van De Weijer, L. Herranz, and S. Jui, "Generalized source-free domain adaptation," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8978-8987, 2021.

[18] H. Li, S.J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5400-5409, 2018.

[19] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5149-5169, 2021.

[20] J. X. Zhang, Z. H. Ling, and L. R. Dai, "Non-parallel sequence-to-sequence voice conversion with disentangled linguistic and speaker representations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 540-552, 2019.

[21] X. Peng, Z. Huang, X. Sun, and K. Saenko, "Domain agnostic learning with disentangled representations," in *Proc. International Conference on Machine Learning (ICML)*, pp. 5102-5112, 2019.

[22] D. Li, Y. Yang, Y. Z. Song, and T. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[23] F. Meneghello, N. Dal Fabbro, D. Garlisi, I. Tinnirello, and M. Rossi, "A CSI dataset for wireless human sensing on 80 MHz Wi-Fi channels," *IEEE Communications Magazine*, vol. 61, no. 9, pp. 146-152, 2023.