Cross-Person Activity Recognition Method Using Snapshot Ensemble Learning

Siyuan Xu†, Zhengran He†, Wenjuan Shi⋆, Yu Wang†, Tomoaki Ohtsuki‡, and Guan Gui†

†College of Telecommunications and Information Engineering, NJUPT, Nanjing, China
⋆College of Physics and Electronic Engineering, Yancheng Teacher University, Yancheng, China
‡Department of Information and Computer Science, Keio University, Yokohama, Japan

***Abstract*—Human activity recognition (HAR) is one of the most promising technologies in the smart home, especially radio frequency (RF-based) method, which has the advantages of low cost, few privacy concerns and wide coverage. In recent years, deep learning (DL) has been introduced into HAR and these DL-based HAR methods usually have outstanding performance. However, as the recognition scenarios and target change, the model performance drops sharply. To solve this problem, we propose a generalized method for cross-person activity recognition (CPAR), which is called snapshot ensemble learning based an attention with bidirectional long short-term memory (SE-ABLSTM). Specifically, by defining the cosine annealing learning rate, the models with diversity are saved and integrated in the same training process. In addition, we provide a dataset for CPAR and simulation results show that our method improves generalization performance by 5% compared to the original method. The source code and dataset for all the experiments can be available at https://github.com/NJUPT-Sivan/Cross-person-HAR.**

***Index Terms*—Human activity recognition, generalization, channel state information, snapshot ensemble.**

I. Introduction

Human activity recognition (HAR) has tremendous applications in the internet of things (IoT) such as context awareness, elderly monitoring and healthcare services [1]–[3]. Typical HAR systems can be divided into three aspects: wearable-based, vision-based and WiFi-based. Wearable-based [4] method has high recognition accuracy but needs to wear additional equipment for activity recognition which is inconvenient. Vision-based [5] approach may pose privacy concerns and is not suitable for deployment in some scenarios. WiFi-based method takes advantage of its characteristics like device-free, wide coverage, low cost and few privacy concerns to win growing concerns in various applications. The signal used for WiFi-based activity recognition can be concluded to two: Received Signal Strength Indication (RSSI) and Channel State Information (CSI). Unlike RSSI, which only reflects the signal strength changes caused by human activities, CSI can obtain both amplitude and phase information from the subcarriers [6]. Human activity can be identified with a high degree of accuracy by analyzing the measured data of CSI, since body movements of different activities can interfere with signal propagation resulting in changes in CSI. Furthermore, human body shapes, speed of performing an activity , environmental obstacles, and other factors will also cause different changes to received CSI signals.

The existing CSI data acquisition tools are mainly rely on Intel 5300 Network Interface Card and Atheros [7], which have high cost and a more complex environment configuration. Deep learning (DL) has been successfully applied in many fields such as intrusion detection [8], signal recognition [9], [10] and wireless communications [11], [12]. Also, DL-based HAR methods can achieve advanced performance. S. Yousefi *et al*. [13], as one of the most cited articles in WiFi sensing, proposed a method based on LSTM network for HAR which can automatically extract information features. Chen *et al*. [14] proposed an attention based bidirectional long short-term memory (ABLSTM) approach that can efficiently recognize various human activities. However, existing WiFI-based methods emphasis more on the recognition accuracy of the trained overall model, ignoring the generalization ability of CPAR. Few work pays attention to domain generalization for HAR [15], [16] and generalization performance is not satisfactory when the model is trained on some people's data and tested on others'. The reason for the decline in the accuracy of CPAR is that there are domain gaps such as age, height, weight, health and other factors, which will make the pace and speed of different people's movements different, resulting in different CSI changes [17].

To improve the generalization performance of CPAR, this paper proposes a SE-ABLSTM method, which uses the cosine annealing learning rate in snapshot ensemble learning [18] to save multiple models with different weights for ensemble prediction in the same training process. The main contributions of this paper are summarized as follows:

- We exploit ESP32 for CSI data collection and public a CSI dataset which has seven different activities including wave, clap, walk, liedown, sitdown, fall and pickup in an indoor environment using ESP32 CSI tool [19].
- Various networks such as CNN, LSTM, BLSTM, and ABLSTM are evaluated for HAR on our dataset and we compare their cross-person generalization performance.
- A SE-ABLSTM method is proposed to improve generalization ability of CPAR. The results show that our method has better generalization ability than the single model and the simple ensemble model.

II. Dataset Description and System Model

As shown in the Fig. 1, the overall system model can be divided into serval parts. The raw CSI data will initially be collected by the ESP32 CSI Tool, and then some necessary preprocessing such as data cleaning will be done. After that we extract features from the CSI data and send the extracted features to the neural network for training. Through the classification and recognition of deep learning neural network, various activities will finally be able to be recognized.

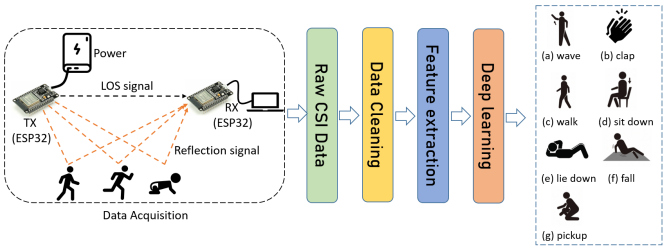


Fig. 1. HAR based on ESP32 CSI Tool.

A. Dataset Description

Our dataset is collected with the ESP32 CSI Tool based on the ESP-IDF framework, and we burn the program through set-target, build, monitor and other operations. We set a higher baud rate of 921600 Baud to achieve a high sampling rate, and the sampling rate is 1000 Hz. CSI data is sent and received through two pieces of ESP32 at a distance of 1.5 meters in an indoor environment. After the two pieces of ESP32 establish a WiFi channel, the CSI corresponding to a specific activity can be collected. A single antenna is used for transmission and reception between the two, and the collected CSI data contains 64 subcarriers, through which we can obtain the amplitude and phase. We perform data cleaning on the raw data to remove those invalid inactivity data and label them with corresponding labels for specific activities.

S. Yousefi *et al*. [13], as one of the most cited papers on WiFi-based HAR, a public dataset of 6 different activities is provided, performed 20 times by 6 users (720 samples). G. Forbes et al. [20], they collected CSI data for 11 activities which were performed 100 times in a home environment (1100 samples). S. Arshad *et al*. [21] collected 720 samples of activities (12 volunteers $\times$ 20 samples $\times$ 3 activities). The author in [22] collect 4 actions with a total of 200-400 samples. In [23], they collected 50 samples for 10 activities (500 samples). F. Moshiri *et al*. [24], they collect 420 samples of activities (3 volunteers $\times$ 20 samples $\times$ 7 activities). Based on other findings, we invited 4 volunteers to perform 7 different activities, including waving, clapping, walking, lying down, sitting, falling, and picking up 20 samples each, resulting in 560 samples. Furthermore, more human activities and different scenes are planned to be collected in the future.

B. System Model

The data acquisition part is to obtain the CSV file containing CSI through the ESP32 CSI Tool, and then process the CSV file to obtain the raw CSI data. Two pieces of ESP32 are used to collect CSI data of various activities, marking the collected data of each action with corresponding tags, and then we perform data cleaning to remove useless activities. The final classification of various activities is obtained by feature extraction and then sent to the neural network for training.

Learning strategies can mainly be divided into machine learning and deep learning. Compared with machine learning such as RF and HMM which limit their wide application due to higher deployment complexity and excessive computing time, deep learning algorithms like CNN and LSTM can extract useful information automatically from CSI which is very helpful for processing and forecasting forecasts for time series series. Sequential models such as RNN and LSTM regard CSI data as a temporal sequence to extract channel features [25].

TABLE I
Comparison of HAR dataset.

Research	Samples	Public Accessibility	CSI Tool
[13]	720	Yes	5300 CSI Tool
[20]	1100	No	Raspberry Pi
[21]	720	No	5300 CSI Tool
[22]	200-400	No	5300 CSI Tool
[23]	500	No	5300 CSI Tool
[24]	420	Yes	Raspberry Pi
Our dataset	560	Yes	ESP32 CSI Tool

III. The Proposed SE-ABLSTM Method

The proposed SE-ABLSTM method consists of two parts, the first part is the structure of the ABLSTM, followed by the snapshot ensemble learning strategy.

A. Network Structure of ABLSTM

Chen *et al*. [14] proposed an attention based bidirectional long short-term memory (ABLSTM) approach for passive HAR which has best accuracy compared to CNN, LSTM and BLSTM. As the Fig. 2 shows that, the raw CSI data is extracted through a bidirectional LSTM network which contains a forward layer and a backward layer, BLSTM can both take past and future CSI information during feature extraction. An attention layer that assigns weights to each feature and time step can automatically learn the importance of each feature and time step. Then the feature matrix trained from the neural network is combined with the weight vector output by the attention layer. In this way, we can finally obtain the model with good robustness and high classification accuracy.

B. Snapshot Ensemble Learning Strategy

An effective snapshot ensemble requires training a neural network with an aggressive learning rate schedulewe define cosine annealing learning rate.

$$\alpha(t) = \frac{\alpha_0}{2}\left(\cos\left(\frac{\pi \mathrm{mod}(t-1, \lceil T/M \rceil)}{\lceil T/M \rceil}\right) + 1\right), \quad (1)$$

where α_0 is the initial learning rate and the value is set to 0.01. Intuitively, T is the total epochs, M is the number of cycles,

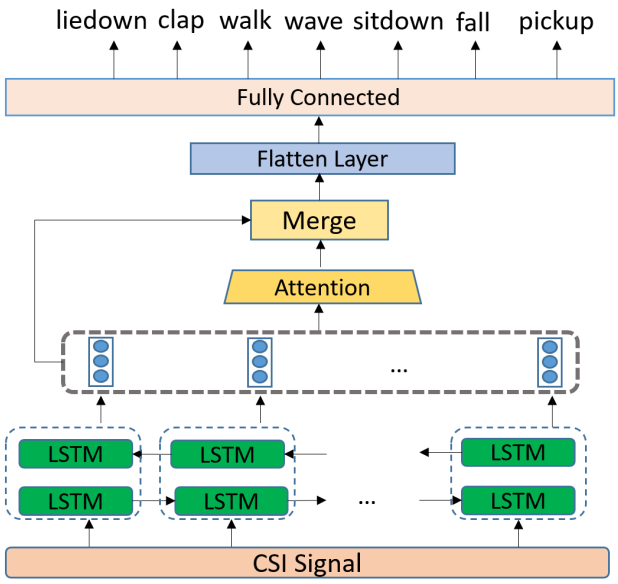


Fig. 2. The structure of ABLSTM.

α_0 falls from 0.01 to 0 representing the process experienced by a cycle.

Algorithm 1 The proposed cross-person activity recognition method using snapshot ensemble learning.

Input: CSI data and corresponding labels in dataset.
Output: Predict accuracy.

Choose and initialize an ABLSTM model.

- α_0: the maximum learning rate.
- T: the total epochs.
- M: the number of cycles.
- $Optimizer$: Adam
- $Loss$: categorical_crossentropy

Define Cosine Annealing Learning Rate:

$$\alpha(t) = \frac{\alpha_0}{2}\left(\cos\left(\frac{\pi \bmod (t-1, \lceil T/M \rceil)}{\lceil T/M \rceil}\right) + 1\right)$$

for $m = 0,\ 1,\ 2,\ \cdots,\ M$ **do**:
  Load the current comprehensive model weight $\omega_{t,0}^{n} = W_t$.
  for $t = 0,\ 1,\ 2,\ \cdots,\ \lceil T/M \rceil$ **do**:
    if learning rate α_t == 0 **then**:
      Save the model and go to the next cycle.
    end if
  end for
  Load all snapshot models for snapshot fusion
  The output of the ensemble is a simple average of the last m models:

$$h_{\text{en}} = \frac{1}{m}\sum_{0}^{m-1} h_i(x).$$

end for
return Predict accuracy

In this paper, we choose to use the Adam optimizer and update the learning rate at each iteration which can improve convergence in the short term even with a larger initial learning rate. We set 10 loops where each loop is set to 50 epochs. Finally, 10 models are trained through one training. The learning rate starts high and decreases relatively quickly to a minimum value near zero before increasing to a maximum value. After creating a custom cosine fire-off function as a callback, models are trained and saved at the bottom of each learning rate schedule. Then, we load the 10 models which have been saved from the previous training and use these models to make ensemble predictions. The CSI data of other untrained people are adopted to test the saved 10 snapshot models, and the accuracy of these tests is average. Compared with the original single model, the generalized performance of the snapshot fusion model is improved. The output of the ensemble h_{en} is a simple average of the last m models.

$$h_{\text{en}} = \frac{1}{m}\sum_{0}^{m-1} h_i(x), \tag{2}$$

where m is the number of the model, x is a test sample and $h_i(x)$ is the softmax score of snapshot i. The trained multiple models are better than the original single model after the integrated learning strategy. As shown in algorithm 1, the cosine fire reduction function is first defined, and then the data of some people is input into the ABLSTM network to obtain multiple models through one training, and use these models to make integrated predictions on the CSI data of others' activities.

IV. Experimental Results

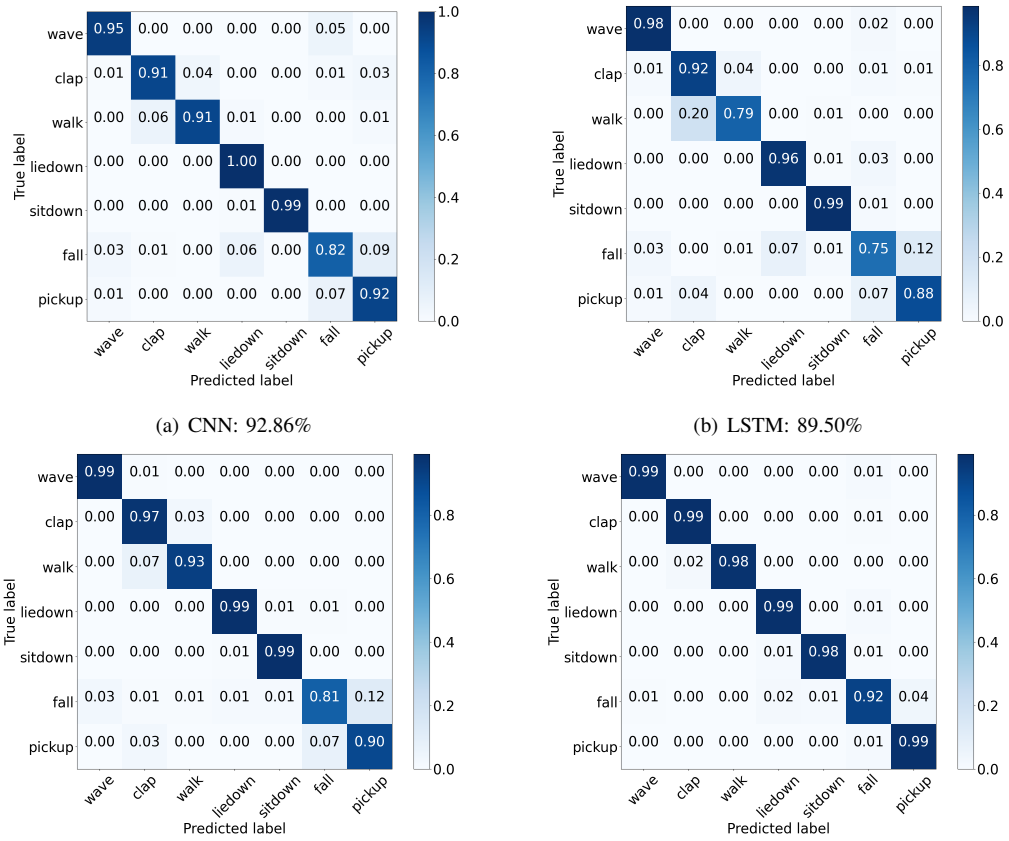
A. Experimental Parameters

The overall experiment can be divided into two tasks : Task I is to use HAR networks to test our dataset and compare the performance differences between these four networks. Task II compares the cross-person generalization performance of different networks and improvement of SE-ABLSTM method proposed in this paper. Simulation is based on Tensorflow as backend, and it is implemented with GeForce GTX 1080 Ti.

- *Task I*: Different networks such as CNN, LSTM, BLSTM and ABLSTM are used to test the performance of our dataset, we divide our dataset into training set, validation set and test set according to the ratio of 8:1:1, and the subsequent simulation results are all test set results.
- *Task II*: Our dataset consists of CSI data of four people with seven different activities. We name the data of these four people as Dataset A, Dataset B, Dataset C, Dataset D. To test the generalization performance of our model, we use three volunteers' CSI data of our dataset to train the model and the fourth to test.

B. Performance of Different Networks in Task I

Networks such as CNN, LSTM, BLSTM and ABLSTM are used to test the performance on the overall dataset. As the Fig. 3 shows, we can see that the accuracy rates of CNN, LSTM, BLSTM, and ABLSTM are respectively 92.86%, 89.50%, 94.01% and 97.48%. ABLSTM has the best performance of these four networks. Regarding the two actions of fall and pickup, due to their similar actions, the recognition rates of CNN, LSTM and BLSTM are not particularly high. In the test on the overall mixed data, ABLSTM has the best recognition accuracy of all.



(a) CNN: 92.86%

(b) LSTM: 89.50%

(c) BLSTM: 94.01%

(d) ABLSTM: 97.48%

Fig. 3. Confusion matrix comparison of different networks.

TABLE II
CROSS-PERSON GENERALIZATION PERFORMANCE OF DIFFERENT NETWORKS.

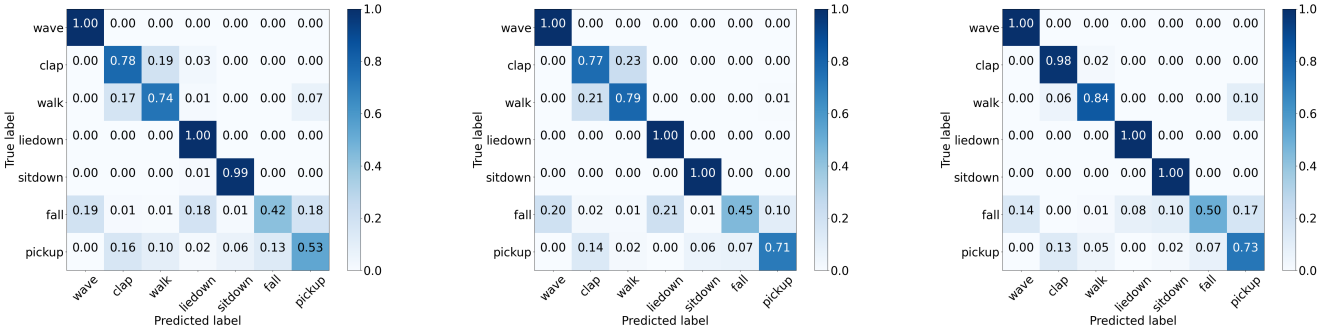
Source	Target	CNN	LSTM	BLSTM	ABLSTM	Simple Ensemble-ABLSTM	SE-ABLSTM (proposed)
{A, B, C}	D	76.86%	68.21%	71.01%	77.29%	78.73%	80.24%
{A, B, D}	C	72.29%	76.43%	72.29%	77.90%	81.71%	86.43%
{A, C, D}	B	77.00%	75.00%	74.14%	79.41%	81.38%	84.87%
{B, C, D}	A	80.14%	85.57%	83.12%	84.96%	89.83%	88.40%
Average		76.57%	76.30%	75.11%	79.89%	82.91%	84.98%

C. Generalization Ability of Different Networks in Task II

All the data of the three people are trained, validated and tested using the four networks of CNN, LSTM, BLSTM and ABLSTM in sequence in the source domain. After that, we will save the obtained model to test the data of the seven activities of the fourth person in the target domain. Finally, we average the four generalization accuracies of each network to get the final generalization performance of the network.

As the Table II shows, among the current four networks, ABLSTM has the relatively best generalization performance. However, the current generalization performance is far from satisfactory. In order to improve the generalization performance of CPAR, we use the strategy of ensemble learning. Ensemble learning integrates a single classifier when the data is classified, and obtains the final classification by combining the classification results of multiple classifiers. To verify the effectiveness of the ensemble learning strategy, we use a simple ensemble learning method to ensemble 10 ABLSTM models. The experimental results show that ensemble learning method helps to improve the generalization performance of the model.

On the basis of ensemble learning, we propose an ABLSTM-based snapshot ensemble learning (SE-ABLSTM) method to further improve the model's generalization performance. We use snapshot ensemble learning method to get 10 models through one training, and then we load these models to perform ensemble prediction after snapshot fusion. From the Fig. 4 we can find out the generalization performance of the four activities of clap, walk, fall, and pickup has been improved. The experimental results show that ensemble learning is effective for improving the generalization ability of the model and snapshot ensemble learning does have better generalization performance than the original single model.



(a) Single ABLSTM: 77.90% (b) Simple ensemble: 81.71% (c) SE-ABLSTM: 86.43%

Fig. 4. Confusion matrix of source {A,B,D} to test target C.

V. Conclusion

In this paper, we propose a SE-ABLSTM method to improve the generalization ability of CPAR. The generalization performance of different networks including CNN, LSTM, BLSTM, and ABLSTM are tested on our dataset. Simulation results show that our snapshot ensemble learning method has better generalization performance compared with the single model and the simple ensemble model. In future work, we try to use the ESP32 CSI Tool to collect more people's CSI activity data to test the generalization performance of the model in the case of large samples, and we will also improve the generalization performance of the model by optimizing the network structure.

References

[1] M. Abdel-Basset, H. Hawash, V. Chang, R. K. Chakrabortty, M. J. Ryan, "Deep learning for heterogeneous human activity recognition in complex IoT applications," *IEEE Internet of Things Jounral*, vol. 9, no. 8, pp. 5653–5665, Aug. 2022.

[2] Z. Xiao, H. Z. Yu, *et al.*, "HarMI: human activity recognition via multi-modality incremental iearning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 939–951, Mar. 2022.

[3] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, J. Cheng, "Data augmentation and dense-LSTM for human activity recognition using WiFi signal," *IEEE Internet of Things Jounral*, vol. 8, no. 6, pp. 4628–4641, Jun. 2021.

[4] L. Oscar, L. Miguel, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, Nov. 2013.

[5] B. Djamila, N. Bini, S. Mohammad, and H. Abdenour, "Vision-based human activity recognition: a survey," *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 30509–30555, Aug. 2020.

[6] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 65–76, Jan. 2011.

[7] S. Souvik, L. Jeongkeun, K. Kyu-Han, and C. Paul, "Avoiding multipath to revive inbuilding WiFi localization," in *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, 2013, pp. 249–262.

[8] S. I. Popoola, R. Ande, *et al.*, "Federated deep learning for zero-day botnet attack detection in IoT edge devices," *IEEE Internet of Things Jounral* , vol. 9, no. 5, pp. 3930–3944, Mar. 2022.

[9] C. B. Hou, G. W. Liu, Q. Tian, Z. C. Zhou, L. J. Hua, and Y. Lin, "Multi-signal modulation classification using sliding window detection and complex convolutional network in frequency domain," *IEEE Internet of Things Jounral*, early access, 2022, doi: 10.1109/JIOT.2022.3167107.

[10] X. X. Zhang, H. T. Zhao, *et al.*, "NAS-AMR: neural architecture search based automatic modulation recognition method for integrating sensing and communication system," *IEEE Transactions on Cognitive Communications and Networking*, early access, 2022, doi: 10.1109/TCCN.2022.3169740.

[11] X. Fu, *et al.*, "Automatic modulation classification based on decentralized learning and ensemble learning," *IEEE Transactions on Vehicular Technology*, early access, 2022, doi: 10.1109/TVT.2022.3164935

[12] G. Gui, J. Wang, J. Yang, M. Liu, and J. L. Sun, "Frequency division duplex massive multiple-input multiple-output downlink channel state information acquisition techniques based on deep learning," *Journal of Data Acquisition and Processing*, vol. 37, no. 3, pp. 502–511, May 2022.

[13] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[14] Z. H. Chen, L. Zhang, *et al.*, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, Oct. 2018.

[15] L. Wang, J. D. Wang, *et al.*, "Local and global alignments for generalizable sensor-based human activity recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 3833–3837.

[16] J. D. Wang, C. L. Lan, *et al.*, "Generalizing to unseen domains: A survey on domain generalization," [Online] available: https://arxiv.org/abs/2103.03097.

[17] H. W. Qian, P. Sinno, *et al.*, "Latent independent excitation for generalizable sensor-based cross-person activity recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 11921–11929.

[18] G. Huang, Y. X. Li, P. Geoff, *et al.*, "Snapshot ensembles: Train 1, get m for free," [Online] available: https://arxiv.org/abs/1704.00109.

[19] S. Hernandez, *et al.*, "Lightweight and standalone IoT based WiFi sensing for active repositioning and mobility," in *A World of Wireless, Mobile and Multimedia Networks*, 2020, pp. 277–286.

[20] G. Forbes, S. Massie, *et al.*, "Wifi-based human activity recognition using Raspberry Pi," in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence*, 2020, pp. 722–730.

[21] S. Arshad, C. H. Feng, Y. H. Liu, *et al.*, "Wi-chase: A WiFi based human activity recognition system for sensorless environments," in *A World of Wireless, Mobile and Multimedia Networks*, 2017, pp. 1–6.

[22] D. Xue, J. Ting, Z. Yi, *et al.*, "Improving WiFi-based human activity recognition with adaptive initial state via one-shot learning," in *IEEE Wireless Communications and Networking Conference (WCNC 2021)*, 2021, pp. 1–6.

[23] Y. Zhang, X. Y. Zhang, *et al.*, "Human activity recognition across scenes and categories based on CSI," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2411–2420, Jul. 2020.

[24] F. Moshiri, S. Reza, N. Mohammad, *et al.*, "A CSI-based human activity recognition using deep learning," *Sensors*, vol. 21, no. 21, pp. 7225–7244, Oct. 2021.

[25] L. bing, C. Wei, W. Wei, *et al.*, "Two-stream convolution augmented transformer for human activity recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 286-293.