

JOINT HUMAN ORIENTATION-ACTIVITY RECOGNITION USING WIFI SIGNALS FOR HUMAN-MACHINE INTERACTION

Hojjat Salehinejad¹, Senior Member, IEEE, Navid Hasanzadeh², Radomir Djogo², and Shahrokh Valaee², Fellow, IEEE

¹Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA

²Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada
hojjat@ieee.org, {navid.hasanzadeh, radomir.djogo}@mail.utoronto.ca, valaee@ece.utoronto.ca

ABSTRACT

WiFi sensing is an important part of the new WiFi 802.11bf standard, which can detect motion and measure distances. In recent years, some machine learning methods have been proposed for human activity recognition from WiFi signals. However, to the best of our knowledge, none of these methods have explored orientation prediction of the user using WiFi signals. Orientation prediction is particularly critical for human-machine interaction in an environment with multiple smart devices. In this paper, we propose a data collection setup and machine learning models for joint human orientation and activity recognition using WiFi signals from a single access point (AP) or multiple APs. The results show feasibility of joint orientation-activity recognition in an indoor environment with a high accuracy.

Index Terms— Activity recognition, channel state information, human-machine interaction, machine learning, WiFi.

1. INTRODUCTION

Human activity recognition (HAR) refers to detection and recognition of human gestures and activities in an environment. Some major systems/mediums for collecting data are wearable sensors (e.g. gyroscope and accelerometer), cameras (e.g. still image and video), and radio frequency signals (e.g. WiFi signals) [1]. HAR with wireless signals has attracted attention due to its privacy preserving nature, broad sensing coverage, and ability to sense the environment without line-of-sight (LoS) [2]. This is particularly interesting since the WiFi 802.11bf standard will enable remote monitoring and sensing [3].

Channel state information (CSI) in a wireless communication system can provide properties about the wireless channel and how a subcarrier has been affected in the environment. Changes in the environment such as walking, falling, and sitting can affect the CSI signals which can be used for various sensing applications. CSI is measured in the baseband and is a vector of complex values. A multiple-input multiple-output (MIMO) wireless system provides a spatial diversity which

can be used for wider and more accurate sensing and detection of activities. This property of wireless signals can be very useful in designing systems for human-machine interaction. Some examples are presence detection [4], security systems [5], localization [6], and internet of things [7].

Various approaches have been proposed for HAR such as long-short term memory (LSTM) [8], sparse auto-encoder (SAE) network [9], attention-based bi-directional LSTM [10], and diversified deep ensemble learning (WiARes) [11]. Most of the proposed methods are based on training many trainable parameters for feature extraction from CSI measurements. This approach requires large CSI training data and hyperparameter tuning. In addition, most of these models due to their high computational complexity may not be suitable for implementation on resource-limited devices such as smart phones and edge devices [12]. LiteHAR [2] method uses a large number of random convolution kernels without training them [13] for feature extraction, followed by a pool of Ridge regression classifiers per frequency for activity recognition. This approach enables fast and accurate HAR using CSI.

To the best of our knowledge, none of the previous works in HAR have explored the possibility of predicting both activity and orientation of the user using CSI. In this paper, machine learning models for prediction of the joint user activity and orientation are introduced. Orientation prediction is particularly important for interaction with devices in smart environments, where multiple devices exist. It governs which device the user is trying to interact with. It is a challenging task, particularly due to the broadcasting and non-LoS properties of the WiFi-based sensing and detection. We have built an infrastructure for CSI measurements collection from multiple access points (APs). Based on our previous work for a lightweight HAR [2] solution, the idea of using 1-dimensional random convolution kernels in [14] is utilized for feature extraction from CSI measurements. Then, Ridge regression classifiers are used for prediction of the activation and orientation of the user. The proposed models are evaluated for single and multiple AP scenarios and the results are discussed.

2. JOINT ORIENTATION-ACTIVITY RECOGNITION MODEL

In this section, we discuss the proposed model for joint human orientation-activity recognition in an indoor environment equipped with one/multiple APs for a single user. First, the feature extraction procedure is introduced. Then, three features classification approaches are proposed.

WiFi CSI amplitude is more stable compared with CSI phase information for HAR, where carrier frequency offset brings random phase errors over different packets [15]. Combined amplitude and phase information can slightly improve localization and HAR accuracy [9] at the cost of increased computational complexity. Let $\mathbf{X}_a \in \mathbb{R}_{\geq 0}^{S \times T}$ represent the CSI amplitudes of AP a with S subcarriers over T indices (i.e. the length of CSI input). For the AP a , the set of N CSI samples is $\{(\mathbf{X}_{a,1}, c_1, o_1), \dots, (\mathbf{X}_{a,N}, c_N, o_N)\}$ where N is the number of samples, c_n is the activity class, and o_n is the orientation class for sample $n \in \{1, \dots, N\}$. In general, the possible orientation and activity classes are finite discrete sets. The set of activity classes is $\mathbf{c} = (c_1, \dots, c_C)$ and the set of orientation classes is $\mathbf{o} = (o_1, \dots, o_O)$, where C is the number of activity classes and O is the number of orientations. The set of samples can be extended for A APs as $\{(\mathbf{X}_{1,1}, \dots, \mathbf{X}_{A,1}, c_1, o_1), \dots, (\mathbf{X}_{1,N}, \dots, \mathbf{X}_{A,N}, c_N, o_N)\}$.

2.1. Feature Extraction

Figure 1(a) shows the feature extraction procedure from a CSI sample \mathbf{X}_n for a single AP. In this approach, based on the multivariate MiniRocket feature extraction method proposed in [14], K 1-dimensional convolution kernels ($\mathbf{w}_1, \dots, \mathbf{w}_K$) are generated where the length of each kernel is fixed and the weights are selected randomly from $\{-1, 2\}$. For each kernel, a set of dilation factors is generated which controls the spread of the kernel over an input with fixed length of T . The set of dilations for kernel k is selected from $\mathcal{L} = \{\lfloor 2^{i \cdot L_{max}/L'} \rfloor | i \in (0, \dots, L')\}$ where L' is a constant, $L_{max} = \log_2((T-1)/(|\mathbf{w}_k| - 1))$ and $L = |\mathcal{L}|$ is the cardinality of \mathcal{L} . This provides $K \times L$ different combinations of kernels and dilations as $\{\mathbf{w}_{k,l} | k \in (1, \dots, K), l \in (1, \dots, L)\}$. The convolution of an input CSI \mathbf{X} with each kernel is

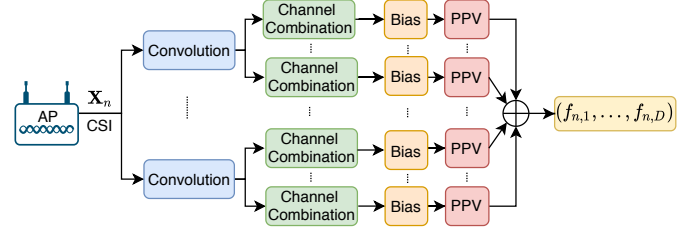
$$\mathbf{u}_{s,k,l} = \mathbf{x}_s * \mathbf{w}_{k,l}, \quad (1)$$

for $s \in (1, \dots, S)$, $k \in (1, \dots, K)$, and $l \in (1, \dots, L)$.

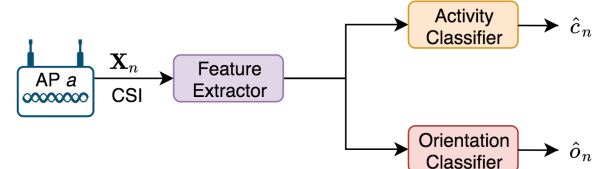
A set of bias terms $\{b_{k,l,j} | j \in (1, \dots, J)\}$ is then calculated based on the quantiles of the convolution output for each pair of kernel and dilation (k, l) . The channel-wise features along with the bias term are then combined as

$$\mathbf{v}_{k,l,j} = \sum_{s=1}^S \mathbf{u}_{s,k,l} - b_{k,l,j}. \quad (2)$$

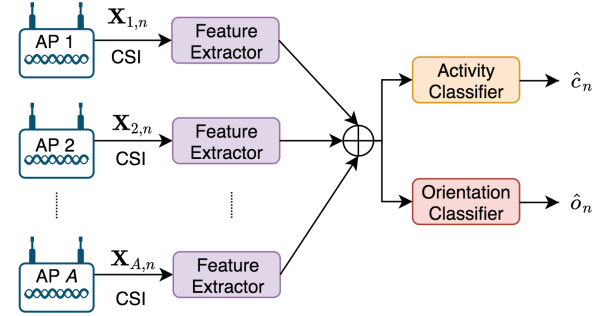
The process of selecting the dilation and bias values is deeply discussed in [14]. It is suggested that the total number of



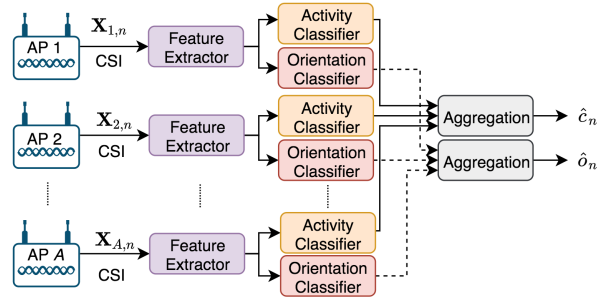
(a) Feature extraction and concatenation (\oplus) for the input channel state information (CSI) \mathbf{X}_n . PPV refers to calculating the portion of positive values using (3). D is the number of extracted features.



(b) Single access point (SAP) model using the feature extractor in (a).



(c) Concatenation of multiple access points (CMAP) model using the feature extractor in (a).



(d) Aggregation of multiple access points (AMAP) model using the feature extractor in (a).

Fig. 1: Proposed feature extraction and joint orientation-activity classification models using a single AP and multiple APs.

extracted features should be kept constant (i.e. $D = 9,996$) as a multiple of K . The number of features can be reduced using feature selection/pruning methods [16]. In [17] a search algorithm is used for reducing D , which can achieve a close performance to the original model. The features are extracted by computing the proportion of positive values (ppv) as

$$f_{k,l,j} = \frac{1}{|\mathbf{v}_{k,l,j}|} \sum_{i=1}^{|\mathbf{v}_{k,l,j}|} \mathbb{1}[v_{k,l,j,i} > 0], \quad (3)$$

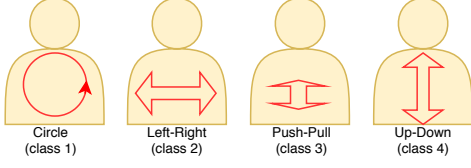


Fig. 2: Four human activities (gestures) used for experiments.

for $k \in (1, \dots, K)$, $l \in (1, \dots, L)$, and $j \in (1, \dots, J_{k,l})$ where $J_{k,l}$ is the number of bias terms and $\mathbb{1}[\cdot]$ is the indicator function. The features can be vectorized for the input CSI signal \mathbf{X}_n as $\mathbf{f}_n = (f_{n,1}, \dots, f_{n,D})$.

2.2. Joint Orientation-Activity Classification

Generally, CSI signals are collected from multiple APs in an indoor environment for HAR applications. In this section, first we introduce an approach for joint orientation-activity recognition from a single AP (SAP) based on the feature extraction procedure discussed in Subsection 2.1. Then, this approach is extended to introduce approaches for aggregation of extracted features from multiple APs (AMAP) and a concatenation of multiple APs (CMAP).

2.2.1. Single Access Point (SAP)

Figure 1(b) shows the setup with a single AP for joint orientation-activity recognition. For a given training dataset, the features \mathbf{f}_n are extracted and passed to two Ridge regression classifiers $\hat{o}_n = \psi_o(\mathbf{f}_n)$ and $\hat{c}_n = \psi_c(\mathbf{f}_n)$, where \hat{o}_n and \hat{c}_n are the predicted orientation class and activation class, respectively, for the input \mathbf{X}_n . This is a general framework and other classifier may be used and evaluated.

2.2.2. Concatenation of Multiple Access Points (CMAP)

A CSI collection setup with multiple APs increases diversity of the signal collection, which enhances sensing of environment. Figure 1(c) shows a setup where A APs are utilized for CSI collection and a feature extractor is implemented per AP. The extracted features are then concatenated as $\mathbf{f}_n = (f_{a,n,d} | a \in (1, \dots, A), d \in (1, \dots, D))$ for each sample. The set of features $\{\mathbf{f}_n | n \in (1, \dots, N)\}$ and the corresponding target classes are then used for training the activity and orientation Ridge regression classifiers.

2.2.3. Aggregation of Multiple Access Points (AMAP)

In the AMAP approach, a feature extractor is allocated per AP followed by a dedicated activity classifier $\hat{c}_{a,n} = \phi_a(\mathbf{f}_{a,n})$ and orientation classifier $\hat{o}_{a,n} = \psi_a(\mathbf{f}_{a,n})$ for $a \in (1, \dots, A)$ and $n \in (1, \dots, N)$. For a given input \mathbf{X}_n , the set of predicted orientations is $\hat{o}_n = (\hat{o}_{a,n} | a \in (1, \dots, A))$ and the set of predicted activities is $\hat{c}_n = (\hat{c}_{a,n} | a \in (1, \dots, A))$. Using

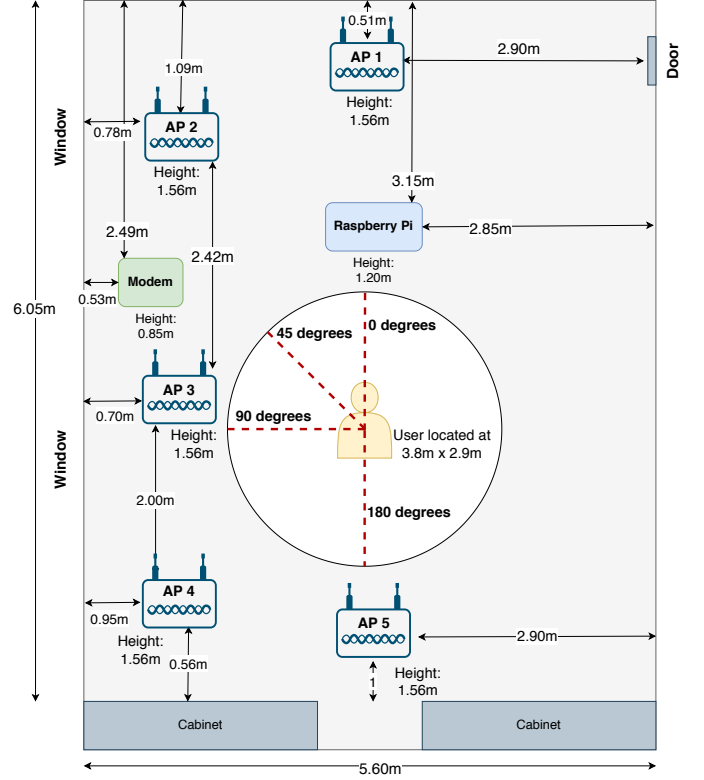


Fig. 3: Detailed floor plan of the data collection setup in an indoor office, which includes location of APs, user, transmitter (Raspberry Pi), and collector (modem).

an aggregation (voting) approach, the predicted activity is

$$\hat{c}_n = \operatorname{argmax}_{c_n \in \mathbf{c}} \left(\sum_{a=1}^A \mathbb{1}[\hat{c}_{a,n}, c_n] \right), \quad (4)$$

and the predicted orientation is

$$\hat{o}_n = \operatorname{argmax}_{o_n \in \mathbf{o}} \left(\sum_{a=1}^A \mathbb{1}[\hat{o}_{a,n}, o_n] \right), \quad (5)$$

where $\mathbb{1}[\hat{i}, i] = 1$ if $\hat{i} = i$ and $\mathbb{1}[\hat{i}, i] = 0$ otherwise.

3. EXPERIMENTS

3.1. Data

We have conducted the experiments for 4 different activity classes (Circle, Left-Right, Push-Pull, Up-Down) as demonstrated in Figure 2. The CSI data was collected at 4 different orientations (0° , 45° , 90° , 180°) as demonstrated in Figure 3. This figure shows our data collection setup which was conducted in an approximately $6m \times 5.6m$ indoor office with 5 APs. The CSI of each AP was read synchronously in a central collector. A Raspberry Pi was used as the transmitter. Per each combination of orientation class and activity class, 20 samples were collected from 6 users. The total number of

Table 1: Classification performance results and standard deviation (in %) over all activity and orientation classes, averaged over 10 independent runs. Acc: Accuracy; BAcc: Balanced accuracy; MCC: Matthews correlation coefficient.

Model	Activity				Orientation			
	Acc	BAcc	F1-Score	MCC	Acc	BAcc	F1-Score	MCC
SAP - AP 1	73.3±1.8	73.3±1.7	73.3±1.8	64.5±2.4	98.0±0.5	98.1±0.5	98.0±0.5	97.4±0.7
SAP - AP 2	69.1±1.4	69.1±1.4	69.1±1.3	58.9±1.8	97.4±0.5	97.4±0.5	97.4±0.5	96.5±0.7
SAP - AP 3	70.3±3.0	70.3±2.9	70.2±3.0	60.4±4.0	98.9±0.5	98.8±0.5	98.9±0.5	98.5±0.7
SAP - AP 4	79.5±1.5	79.6±1.5	79.5±1.5	72.7±2.0	98.7±0.5	98.7±0.5	98.7±0.5	98.2±0.7
SAP - AP 5	82.7±1.5	82.8±1.5	82.7±1.5	77.0±2.0	99.4±0.4	99.4±0.4	99.4±0.4	99.2±0.6
AMAP	91.1±1.8	91.1±1.8	91.1±1.8	88.1±2.4	99.0±0.1	99.0±0.1	99.0±0.1	99.0±0.1
CMAF	91.4±1.4	91.4±1.5	91.4±1.5	88.5±1.9	99.7±0.2	99.7±0.2	99.7±0.2	99.6±0.2

Table 2: Classification accuracy results and standard deviation (in %) per activity class and orientation class, averaged over 10 independent runs.

Model	Activity				Orientation			
	Circle	Left-Right	Push-Pull	Up-Down	0°	45°	90°	180°
SAP - AP 1	78.6±3.1	71.1±5.5	77.3±2.0	66.0±4.8	98.5±1.3	98.3±1.1	98.1±0.6	97.4±1.7
SAP - AP 2	72.9±3.1	65.7±5.6	71.0±3.2	66.9±3.3	96.7±1.5	98.0±1.1	98.0±1.5	96.8±1.5
SAP - AP 3	77.0±4.1	67.9±4.1	71.9±6.6	64.4±3.2	99.3±0.6	99.5±0.6	97.4±1.2	99.1±1.1
SAP - AP 4	81.3±4.2	78.6±2.1	82.5±2.5	75.9±5.1	99.5±0.7	98.5±1.1	97.7±1.2	99.0±1.7
SAP - AP 5	81.7±3.0	81.0±3.8	84.5±2.9	83.8±3.2	99.4±0.5	99.2±0.7	99.6±0.6	99.6±0.4
AMAP	92.4±2.6	86.0±2.6	94.9±2.6	84.1±2.7	99.0±0.1	99.0±0.1	99.0±0.1	99.0±0.1
CMAF	93.6±2.4	89.5±3.2	93.4±2.6	88.9±3.8	99.6±0.7	99.7±0.4	99.9±0.3	99.8±0.4

collected samples from each AP was $20 \times 4 \times 4 \times 6 = 1,920$, where 80% was used for training and 20% was used for testing the models.

3.2. Setup

The classifiers were cross-validated with (0.001, 0.01, 0.1, 1) regularization strengths. The results are averaged over 10 independent runs. The PyTorch implementation¹ of MiniRocket [14] with a fixed set of $K = 84$ kernels of length 9 and the total number of features of $D = 9,996$. The models were implemented in PyTorch and were trained on a single NVIDIA GTX GPU.

3.3. Classification Performance Analysis

Classification performance of the SAP, AMAP, and CMAF models with respect to the accuracy (Acc), balanced accuracy (BAcc), F1-Score, and Matthews correlation coefficient (MCC) metrics is presented in Tables 1 and 2. In Table 1, the average performance results over all activity and orientation classes are presented. The SAP model was trained and evaluated per each AP independently. The results show that the SAP model with AP 5 has a better performance than the other SAP models for both activity and orientation recognition tasks. As Figure 3 shows, the user is located between the shortest path between the AP and the transmitter. However, the other APs have a shortest LoS with the transmitter without direct interference with the user. Hence, proper placement of the APs with respect to the user and transmitter location can improve sensing of the environment and achieving a higher activity and orientation recognition accuracy.

¹<https://github.com/timeseriesAI/tsai/blob/main/tsai>

The overall results show that the CMAF and AMAP approaches have a competitive performance, better than the SAP evaluations. CMAF performs slightly better than AMAP in activity recognition but has a lower performance in orientation prediction. All approaches, even with a single AP, have a high performance in prediction of the orientation of the user. This is particularly important in recognizing which device/orientation a user is interacting with. Granular performance results per activity class and orientation class in Table 2 show that the *Up-Down* activity is relatively more challenging to recognize than the other gestures. The performance per orientation class is high for all approaches and the performance difference between different classes is not significant. Overall, the CMAF approach has a relatively better performance and less complexity due to using a single joint activity and orientation classifier.

4. CONCLUSIONS

In this paper, we explore joint prediction of human's orientation and activity using WiFi signals for human-machine interaction in indoor environments. Models based on random convolution kernels without training them are proposed for feature extraction followed by Ridge regression classifier, which can be deployed on resource-limited devices. Our results show that increasing the spatial diversity of WiFi signal collection by utilizing multiple APs can increase the classification accuracy of human activities. However, it is possible to predict orientation of the user using a single AP with a high accuracy. A cross-domain evaluation can be conducted in future works.

5. REFERENCES

- [1] Fuqiang Gu, Mu-Huan Chung, Mark Chignell, Shahrokh Valaee, Baoding Zhou, and Xue Liu, "A survey on deep learning for human activity recognition," *ACM Computing Surveys (CSUR)*, vol. 54, no. 8, pp. 1–34, 2021.
- [2] Hojjat Salehinejad and Shahrokh Valaee, "Litehar: Lightweight human activity recognition from wifi signals with random convolution kernels," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 4068–4072.
- [3] Francesco Restuccia, "Ieee 802.11 bf: Toward ubiquitous wi-fi sensing," *arXiv preprint arXiv:2103.14918*, 2021.
- [4] Simone Di Domenico, Mauro De Sanctis, Ernestina Cianca, and Marina Ruggieri, "Wifi-based through-the-wall presence detection of stationary and moving humans analyzing the doppler spectrum," *IEEE Aerospace and Electronic Systems Magazine*, vol. 33, no. 5-6, pp. 14–19, 2018.
- [5] Shaohu Zhang, Raghav H Venkatnarayan, and Muhammad Shahzad, "A wifi-based home security system," in *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 2020, pp. 129–137.
- [6] Hojjat Salehinejad, Robert Zadeh, Ramiro Liscano, and Shahryar Rahnamayan, "3d localization in large-scale wireless sensor networks: A micro-differential evolution approach," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*. IEEE, 2014, pp. 1824–1828.
- [7] Pritam Khan, Bathula Shiva Karthik Reddy, Ankur Pandey, Sudhir Kumar, and Moustafa Youssef, "Differential channel-state-information-based human activity recognition in iot networks," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11290–11302, 2020.
- [8] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon, and Shahrokh Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.
- [9] Qinhua Gao, Jie Wang, Xiaorui Ma, Xueyan Feng, and Hongyu Wang, "Csi-based device-free wireless localization and activity recognition using radio image features," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10346–10356, 2017.
- [10] Zhenghua Chen, Le Zhang, Chaoyang Jiang, Zhiguang Cao, and Wei Cui, "Wifi csi based passive human activity recognition using attention based blstm," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.
- [11] Wei Cui, Bing Li, Le Zhang, and Zhenghua Chen, "Device-free single-user activity recognition using diversified deep ensemble learning," *Applied Soft Computing*, vol. 102, pp. 107066, 2021.
- [12] Hojjat Salehinejad and Shahrokh Valaee, "Edropout: Energy-based dropout and pruning of deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5279–5292, 2022.
- [13] Angus Dempster, François Petitjean, and Geoffrey I Webb, "Rocket: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1454–1495, 2020.
- [14] Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb, "Minirocket: A very fast (almost) deterministic transform for time series classification," *arXiv preprint arXiv:2012.08791*, 2020.
- [15] Harshit Ambalkar, Xuyu Wang, and Shiwen Mao, "Adversarial human activity recognition using wi-fi csi," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 2021, pp. 1–5.
- [16] Hojjat Salehinejad and Shahrokh Valaee, "A framework for pruning deep neural networks using energy-based models," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3920–3924.
- [17] Hojjat Salehinejad, Yang Wang, Yuanhao Yu, Tang Jin, and Shahrokh Valaee, "S-rocket: Selective random convolution kernels for time series classification," *arXiv preprint arXiv:2203.03445*, 2022.