Direction-Independent Human Activity Recognition Using a Distributed MIMO Radar System and Deep Learning

Sahil Waqar, Muhammad Muaaz, and Matthias Pätzold, *Senior Member, IEEE*

Abstract—**Modern monostatic radar-based human activity recognition (HAR) systems perform very well as long as the direction of human activities is either toward or away from the radar. The monostatic single-input–single-output (SISO) and monostatic multiple-input–multiple-output (MIMO) radar systems cannot detect motion of an object that moves perpendicularly to the radar's boresight axis. Due to this physical layer limitation, today's radar-based HAR systems fail to classify multidirectional human activities. In this article, we resolve this typical but critical physical layer problem of contemporary HAR systems. We propose a HAR system underlying a distributed MIMO radar configuration, where multiple antennas of a millimeter wave (mm-wave) MIMO radar system (Ancortek SDR-KIT 2400T2R4) are distributed in an indoor environment. In our proposed HAR system, we have two independent and identical monostatic radar subsystems that irradiate and capture the multidirectional human movement from two perspectives, which allows to compute two distinct time-variant (TV) radial velocity distributions. A feature extraction network extracts numerous features from the measured TV radial velocity distributions, which are then fused by a multiclass classifier to detect five types of human activities. The proposed multiperspective MIMO-radar-based HAR system achieves a classification accuracy of 98.52%, which surpasses the accuracy of SISO radar-based HAR system by more than 9%. Our approach resolves the physical layer limitations of modern HAR systems that are based on either monostatic SISO or monostatic MIMO radar systems.**

Index Terms—**Deep learning, direction-independent human activity recognition (HAR), fall detection, feature fusion, multistatic radar, multiview radar sensing, orientation-independent HAR.**
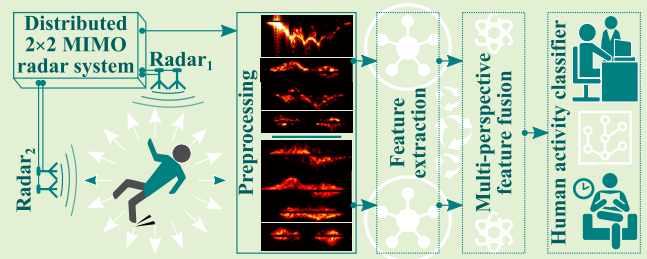
I. INTRODUCTION

A. General Background

STUDIES have shown a considerable amount of progress in the area of human activity recognition (HAR) over the past few years [1], [2], [3], [4]. The steady interest in HAR is due to its extensive range of applications. Over the years, HAR systems have proven their usefulness in application areas such as social robotics [5], autonomous driving [6], sports [7], [8], home automation [9], healthcare [10], automated video analysis [11], and human–computer interaction [12].

To date, numerous diverse sensing modalities have been adopted to effectuate the HAR task. However, each modality may exhibit its distinct advantages and disadvantages [13]. For instance, due to the ongoing advancements in computer vision techniques, HAR systems based on vision sensors have produced remarkable results [14], [15]. However, vision sensors are often criticized because they are very susceptible to lighting conditions, occlusion, and can violate user privacy. Wearable sensors [16], [17], [18], [19] on the other hand, despite being very effective HAR sensors, are generally criticized for being fragile, obtrusive, and vulnerable to user negligence. Also, the need to be worn indefinitely renders the wearable sensors impractical and inconvenient, especially for elderlies or infirmed persons. By taking into consideration the aforementioned shortcomings, recently HAR systems based on radio frequency (RF) sensing techniques have been preferred more and more despite new challenges and hurdles.

Lately, many researchers have studied and eventually leaned toward Wi-Fi and radar systems for the HAR purpose [20], [21], [22], [23]. Unlike radar systems, commercial grade Wi-Fi routers have the channel frequency response with notably noisy phases [24], [25], [26], [27]. In contrast, commercial coherent radar systems conserve the phase information within

their coherent processing interval (CPI) [28]. Thus, small phase variations corresponding to nonstationary scatterers in an environment can be easily processed by coherent signal processing techniques [29], [30]. This is one of the reasons why coherent radar systems have been preferred over Wi-Fi devices to capture the propagation phenomena caused by complex human activities. In the context of RF sensing, the recognition of human activity often relies on exploiting the micro-Doppler phenomenon [31], [32], [33], [34] to discern the specific type of activity being performed. Thanks to recent advancements in the areas of radar techniques and machine/deep learning, the classification and tracking of a wide range of human activities in complex environments will be within reach in a few years.

B. Problem Description

A major problem of radar-based HAR systems is their inability to generate an adequate micro-Doppler signature in a situation where a person moves perpendicularly to the radar boresight axis. Our article is a step in this direction, in which we propose a pragmatic solution to the problem of direction-independent HAR. Thus, we will look into the classification of five different types of human activities performed in different directions. Single monostatic radar-based HAR systems do not consider the direction of human motion and thus tend to fail in classifying human activities performed in different directions.

Erol and Amin [35] reported the average classification performance at different aspect angles for a human falling activity. For a human fall parallel to the radar boresight axis or at 0° aspect angle, the classification accuracy was 96%; at 60° aspect angle, the classification accuracy dropped to 85%; and at 90° aspect angle (falling perpendicularly to the radar boresight axis), the classification accuracy plummeted to 45% rendering the HAR system futile. Similarly, for six human activities, Ding et al. [36] reported a decrease in classification performance from 95.8% to 86.7% by changing the radar's viewing angle from 15° to 30°.

C. Related Work

Some of the approaches to mitigate the problem of the direction of human motion are discussed here along with their shortcomings. In [37] and [38] it was shown that by positioning a radar on the ceiling, a human falling in different directions can be detected, but the solution cannot be generalized to classify more complex human activities. To realize a direction-independent HAR, it is tempting to employ a monostatic beamforming multiple-input–multiple-output (MIMO) radar system with the capability of measuring the target's angle [39], [40]. But in practice, commercial beamforming radar systems have poor angular and cross-range resolutions due to their limited hardware resources. Thus, for applications such as short-range hand gesture sensing, where the cross-range resolution is not a concern, Molchanov et al. [41] rightly utilized the angular information of a single-input–multiple-output (SIMO) frequency-modulated continuous wave (FMCW) monopulse radar. Unfortunately, the approach cannot be extended to direction-independent HAR systems because of the radar's poor cross-range resolution. Recently, HAR systems are realized by using three-dimensional (3-D) point cloud data

generated by millimeter wave (mm-wave) monostatic MIMO radar systems [42], [43]. But 3-D point cloud data also suffer from the problem of poor cross-range resolution. For better angle estimation or, equivalently, cross-range resolution, more advanced signal processing techniques such as the *"estimation of signal parameters via rotational invariance techniques (ESPRIT)"* [44] and *"multiple signal classification (MUSIC)"* algorithms [45], [46] are usually employed, but these estimation techniques demand a high signal-to-noise ratio [47]. Alternatively, a single-input–single-output (SISO) bistatic radar system [48] is a good choice for HAR. However, an even better choice for the direction-independent HAR is multiperspective multistatic MIMO radar systems. They can provide the best multiview signatures of human activities, as we will see in this article.

D. Proposed Approach for HAR

To overcome the aforementioned issues and drawbacks of monostatic SISO, SIMO, and beamforming MIMO radar-based HAR systems, we develop a multiperspective 2×2 distributed MIMO radar system to realize a direction-independent HAR system. In our approach, two radar subsystems, each consisting of one transmit and one receive antenna and their own independent signal preprocessing units, are spatially distributed to irradiate the environment from different perspectives (see Section III). This multistatic MIMO radar framework enables us to detect and classify different types of human activities independent of their respective directions.

Human body segments can be modeled by N moving scatterers, which reflect back the radar signals to the radar receiver. The scatterers' distinct time-variant (TV) radial velocity components can be described by the so-called TV radial velocity distribution (see Section II). The TV radial velocity distributions at the output of the radar's signal preprocessor are in fact the input feature maps to our classifier, which is based on a deep convolutional neural network (DCNN). We use deep learning methods to automatically extract the features from the TV radial velocity distributions of the MIMO radar system to finally classify the type of human activity regardless of its direction of motion.

Conventionally, it was not uncommon to manually extract features in single-variable and joint-variable domains to classify human activities using machine learning techniques, such as support vector machine (SVM), with a well-documented classification accuracy of 90% [49]. Widely adopted conventional machine learning algorithms in conjunction with domain-based feature engineering usually have theoretical foundations and are computationally less expensive when compared to deep learning algorithms. However, manual feature engineering is quite cumbersome and requires specific expertise. Determining the relevance and significance of features for identifying specific motion artifacts is also a complicated task. Large differences in manually measured features were found in different individuals monitored for health status, body height, and habits [50]. Therefore, to account for the intricate attributes of human motion, and to overcome the aforementioned challenges associated with manual feature engineering, deep learning algorithms are preferred [51].

To train and test our SISO and MIMO radar-based direction-independent HAR classifiers (see Section V), we recorded a novel HAR dataset, where the human activities were performed in several directions in the two-dimensional (2-D) horizontal xy plane. In this regard, we denote the recorded HAR dataset with the superscript "(2-D)" as HAR[(2-D)] (see Section IV). For a conventional monostatic SISO radar-based HAR classifier that contains the human movement merely along the one-dimensional (1-D) x-axis or the monostatic radar's boresight, we denote the recorded HAR dataset accordingly by HAR[(1-D)].

E. Contributions

The MIMO radar-based HAR system presented in this article is a stride forward toward actualizing more advanced RF-based HAR systems. The main contributions of the research are as follows.

1) For our direction-independent HAR system, we have addressed a critical physical layer problem of monostatic radar systems related to the target's aspect angle.

2) For a monostatic SISO and the multistatic MIMO radar configurations, we have analyzed the variations in measured channel characteristics for five types of human activities (falling, walking, standing, sitting, picking). We also studied the effects of different directions of human activities by analyzing the TV radial velocity distributions of the MIMO radar system (see Section III).

3) We composed a completely novel HAR dataset, denoted as HAR[(2-D)], by using the multiperspective 2×2 MIMO radar configuration (see Section IV). We recorded real human activities by using a commercial mm-wave radar system known as Ancortek SDR-KIT 2400T2R4. The HAR[(2-D)] dataset consists of five types of human activities performed by six different persons in several directions.

4) By using the HAR[(2-D)] dataset and its derivative or subset dataset denoted as HAR[(1-D)], we have developed and analyzed three different HAR systems (see Section V): 1) a SISO radar-based conventional HAR system; 2) a SISO radar-based direction-independent HAR system; and 3) a MIMO radar-based direction-independent HAR system. The proposed 2×2 MIMO radar-based HAR system is capable of recognizing human gross motor activities regardless of the aspect angle or direction of motion, and it is straightforwardly scalable to a higher number of antennas for a more complex human activity classification task.

5) For the three HAR systems, we accordingly designed three different DCNN-based multiclass classifiers. The DCNN classifier extract features automatically from the radar's TV radial velocity distribution before classifying an activity. For the distributed MIMO radar-based classifier, feature level fusion has been adopted, which virtually combines the target's information from different aspect angles, and thereby eradicates the limitations that emerge due to the direction of motion.

6) The classification performances of the three HAR systems have been assessed and compared quantitatively.

It is shown that the proposed HAR system, based on the multiperspective 2×2 MIMO radar framework, improves the classification accuracy of the monostatic SISO radar-based HAR system from 88.98% to 98.52%.

F. Article Organization

The article organization is as follows. Section II describes the MIMO radar system model and the deep learning methods that are utilized in this research. A critical problem of modern SISO and monostatic MIMO radar-based HAR systems and its solution is discussed in Section III. The data acquisition campaign is described in Section IV. In Section V, a conventional and a direction-independent SISO radar-based HAR system, as well as a direction-independent MIMO radar-based HAR system are presented. Lastly, Section VI draws the conclusions.

II. SYSTEM OVERVIEW

A. MIMO Radar Signal Preprocessing

An FMCW 2×2 MIMO radar system periodically transmits a chirp waveform $c_i(t')$, which can be expressed as [52]

$$c_i(t') = \exp\left[j\left(\phi_i + 2\pi f_0 t' + \gamma \pi t'^2 \right) \right], \quad 0 \leq t' < T_{\text{sw}} \quad (1)$$

where $i = 1, 2$. The symbol ϕ_i is the initial phase term, f_0 is the initial frequency, and γ is the slope of the chirp waveform. The symbols t' and T_{sw} in (1) are the fast time and duration of the chirp, respectively. We adopted a time division multiple access (TDMA) scheme, where the transmitter antenna $A_i^{T_x}$ periodically transmits the chirp waveform $c_i(t')$ in separate time windows, which are defined as $(2n + i - 1)T_{\text{sw}} \leq t' < (2n+i)T_{\text{sw}}$ for $n = 0, 1, \ldots$ and $i = 1, 2$. With the help of the Dirac delta function $\delta(\cdot)$, we can express the transmit signal $s_i(t', t)$ in terms of fast time t' and slow time t as [53]

$$s_i(t', t) = \sum_{n=0}^{\infty} c_i(t')\delta(t - T_{n,i}). \quad (2)$$

The symbol $T_{n,i}$ in (2) is the discrete slow time that depends on the chirp duration T_{sw} according to $T_{n,i} = (2n+i-1)T_{\text{sw}}$.

For a 2×2 MIMO radar, the notation $A_i^{T_x}$–$A_k^{R_x}$ describes the wireless link between the transmitter antenna $A_i^{T_x}$ and the receiver antenna $A_k^{R_x}$. The transmit signal $s_i(t', t)$ interacts with $\mathcal{L}$ stationary and nonstationary scatterers present in the wireless link $A_i^{T_x}$–$A_k^{R_x}$, where $i, k \in \{1, 2\}$. Let the symbols $d_{ik}^{(l)}$, c_0, and λ denote the propagation distance of the lth scatterer, speed of light, and radar's wavelength, respectively. Then, the beat frequency $f_{b,ik}^{(l)}$ and the phase $\phi_{ik}^{(l)}$ of the lth scatterer are given by $f_{b,ik}^{(l)} = 2d_{ik}^{(l)}\gamma/c_0$ and $\phi_{ik}^{(l)} = 4\pi d_{ik}^{(l)}/\lambda$, respectively, where $l = 1, 2, \ldots, \mathcal{L}$. For the wireless link $A_i^{T_x}$–$A_k^{R_x}$ and the lth scatterer, the received beat signal $s_{b,ik}^{(l)}(t', t)$ can be expressed as [53]

$$s_{b,ik}^{(l)}(t', t) = \sum_{n=0}^{\infty} a_{ik}^{(l)} \exp\left[j\left(2\pi f_{b,ik}^{(l)} t' + \phi_{ik}^{(l)} \right) \right]\delta\left(t - T_{n,i} - \tau_{ik}^{(l)} \right)$$

$$(3)$$

where $a_{ik}^{(l)}$ is the gain, which is assumed to be constant within the radar's CPI. The propagation delay $\tau_{ik}^{(l)}$ in (3) is related to

the beat frequency $f_{b,ik}^{(l)}$ by $\tau_{ik}^{(l)} = f_{b,ik}^{(l)}/\gamma$. At the radar receiver, the composite beat signal $s_{b,ik}(t', t)$ is simply the sum of all $\mathcal{L}$ beat signals, i.e.,

$$s_{b,ik}(t', t) = \sum_{l=1}^{\mathcal{L}} s_{b,ik}^{(l)}(t', t). \tag{4}$$

We obtain the beat frequency function $S_{b,ik}(f_b, t)$ by computing the Fourier transform of the beat signal $s_{b,ik}(t', t)$ over the fast time t', i.e., [54]

$$S_{b,ik}(f_b, t) = \int_0^{T_{sw}} s_{b,ik}(t', t)e^{-j2\pi f_b t'} dt' \tag{5}$$

where f_b is the beat frequency. The beat frequency function $S_{b,ik}(f_b, t)$ in (5) further undergoes a short-time Fourier transform (STFT) over the slow time t. Subsequently, the square of the STFT results in the TV micro-Doppler signature $S_{ik}(f, t)$, which is given as

$$S_{ik}(f, t) = \left| \int_0^{f_{b,max}} \int_{-\infty}^{\infty} S_{b,ik}(f_b, t'')W_r(t'' - t)e^{-j2\pi f t''} dt'' df_b \right|^2 \tag{6}$$

where f represents the Doppler frequency, $f_{b,max}$ is the maximum beat frequency, t'' is the running time, and $W_r(\cdot)$ is a window function, which is in our case a rectangular function with a width of $64T_{sw}$.

Finally, the TV radial velocity distribution $p_{ik}(v, t)$ is obtained from the TV micro-Doppler signature $S_{ik}(f, t)$ according to [53]

$$p_{ik}(v, t) = \frac{S_{ik}\left(\frac{2f_0}{c_0}v, t\right)}{\int_{-\infty}^{\infty} S_{ik}\left(\frac{2f_0}{c_0}v, t\right) dv} \tag{7}$$

where v represents the radial velocity. Note that the human body is composed of body segments and each body segment contains several scatterers that reflect back the RF signals to the radar. Each scatterer on a human body segment has a unique TV radial velocity component due to its spatially distinct motion. The TV radial velocity distribution $p_{ik}(v, t)$ contains the radial velocity components from all the scatterers on the human body. We use the expression in (7) to obtain the TV radial velocity distribution $p_{ik}(v, t)$ of the recorded human activities. The TV radial velocity distribution $p_{ik}(v, t)$ is converted into an image in the time-velocity domain, which is basically an input feature map to the DCNN, as described in Section II-B.

B. Deep Learning

In this section, a supervised learning-based multiclass classification method is delineated. Assume a d-dimensional mth feature vector $\boldsymbol{x}_m$ that belongs to a feature space $\mathcal{X}$. This feature space $\mathcal{X}$ is a proper subset of the real coordinate space $\mathbb{R}^d$, meaning that $\boldsymbol{x}_m \in \mathcal{X} \subset \mathbb{R}^d$. For the entire number of classes C, the mth label y_m is an element of a label space $\mathcal{Y} = \{1, 2, \ldots, C\}$, i.e., $y_m \in \mathcal{Y}$. A dataset D is defined as $\{(\boldsymbol{x}_m, y_m)\}_{m=0}^{M-1}$, where M is the total number of labeled training samples.

We aim to design a DCNN-based classifier function $\mathcal{C}_f$ that maps the input feature space $\mathcal{X}$ into the label space $\mathcal{Y}$, i.e., $\mathcal{C}_f : \mathcal{X} \mapsto \mathcal{Y}$. An empirical risk $R_J(\mathcal{C}_f)$ corresponding to the categorical cross-entropy loss function J_{CCE} and the classifier function $\mathcal{C}_f$ is given as [55], [56]

$$R_J(\mathcal{C}_f) = \mathbb{E}_D\{J_{CCE}(\mathcal{C}_f(\boldsymbol{x}; \boldsymbol{\theta}), y_{\boldsymbol{x}})\}$$
$$= -\frac{1}{M} \sum_{m=0}^{M-1} \sum_{c=0}^{C-1} y_m^c \log \mathcal{C}_f^c(\boldsymbol{x}_m; \boldsymbol{\theta}) \tag{8}$$

where $\mathbb{E}_D\{\cdot\}$ denotes the expectation operator that is performed over the empirical distribution, which can either be the dataset D or a mini-batch from the dataset D. In (8), the symbol $\boldsymbol{\theta}$ is a vector of trainable parameters defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)$, where L depends on the complexity of the classifier. The symbol y_m^c in (8) is the cth entity of the mth one-hot encoded label vector $\boldsymbol{y}_m$, which means $y_m^c \in \{0, 1\}$ such that $(\mathbf{1})^\top \boldsymbol{y}_m = 1 \ \forall \ m$, where $\mathbf{1}$ is a C-dimensional vector of ones, and $(\cdot)^\top$ is the transpose operator. The symbol $\mathcal{C}_f^c$ represents the cth element of the classifier function $\mathcal{C}_f$. We have used the softmax layer as an output layer of the deep neural network (DNN), thus $\sum_{c=0}^{C-1} \mathcal{C}_f^c(\boldsymbol{x}_m; \boldsymbol{\theta}) = 1$, and $\mathcal{C}_f^c(\boldsymbol{x}_m; \boldsymbol{\theta}) \geq 0, \ \forall \ m, c, \boldsymbol{\theta}$. The trainable parameters of the vector $\boldsymbol{\theta}$ corresponding to the classifier function $\mathcal{C}_f$ can be obtained by minimizing the empirical risk $R_J(\mathcal{C}_f)$.

The learning process of the DCNN and DNN is the same, but in case of DCNN, the number of trainable parameters is drastically reduced. In a DCNN, convolutional layers are employed to generate the feature maps from their inputs by means of multiple learnable filters. Assume a total number of Q filters in a convolutional layer, then the mth input feature map $\boldsymbol{x}_m$ is convolved with the qth filter. The qth filter is characterized by its trainable weight vector $\mathbf{w}_q$ and bias b_q. Then, the qth output $\mathbf{y}_q$ of the convolutional layer is given by

$$\mathbf{y}_q = \sum_{m=1}^{M} \sigma(\boldsymbol{x}_m * \mathbf{w}_q + b_q \mathbf{1}), \quad q = 1, 2, \ldots, Q \tag{9}$$

where the symbol $*$ denotes the convolutional operator. The function $\sigma(\cdot)$ in (9) is a rectified linear unit (ReLU) activation function [57] formulated as $\sigma(x) = \max(0, x)$, which mitigates the problems of slow convergence and gradient vanishing [58].

Pooling layers are generally utilized as an abstraction and downsampling tool to progressively reduce the spatial size and redundancies of the extracted feature maps to increase the network's computational efficiency. Moreover, dropout layers are added to the network to improve the network generalizability and to avoid the overfitting problem [59]. After several convolutional layers, the feature maps are flattened before feeding them to the fully connected dense layers or multilayer perceptron (MLP) layers.

In this research, we use a stochastic optimization technique known as adaptive moment estimation (Adam) [60] to optimize or train the parameters of the vector $\boldsymbol{\theta}$. The Adam algorithm applies adaptive learning rates that are based on

the estimates of the first-order moment $\boldsymbol{m}_\kappa$ and second-order moment $\boldsymbol{v}_\kappa$ of the gradient $\boldsymbol{g}_\kappa$ according to

$$\boldsymbol{m}_\kappa = \beta_1 \boldsymbol{m}_{\kappa-1} + (1-\beta_1)\boldsymbol{g}_\kappa \qquad (10)$$

and

$$\boldsymbol{v}_\kappa = \beta_2 \boldsymbol{v}_{\kappa-1} + (1-\beta_2)\boldsymbol{g}_\kappa^2 \qquad (11)$$

where the symbol κ denotes the iteration number, and the decay factors are denoted by β_1 and β_2. The gradient $\boldsymbol{g}_\kappa$ in (10) and (11) is the gradient of the stochastic objective function $f(\boldsymbol{\theta}_\kappa) = \min_{\boldsymbol{\theta}_\kappa} R_J(\mathcal{C}_f)$. Note that in the Adam algorithm, element-wise operations are adopted for all the vectors $\boldsymbol{m}_\kappa$, $\boldsymbol{v}_\kappa$, $\boldsymbol{g}_\kappa$, and $\boldsymbol{\theta}_\kappa$. Additionally, to counteract the initialization bias of the moments or to avoid the moments' biasedness toward zero, Kingma and Ba [60] suggested that the first- and second-order moments can be rectified as $\hat{\boldsymbol{m}}_\kappa = \boldsymbol{m}_\kappa/(1-\beta_1^\kappa)$ and $\hat{\boldsymbol{v}}_\kappa = \boldsymbol{v}_\kappa/(1-\beta_2^\kappa)$, respectively. Then, for α_κ being the learning rate and ϵ a small constant, the ℓth parameter of the vector $\boldsymbol{\theta}_\kappa$ at the κth iteration can be updated as [60]

$$\theta_{\ell,\kappa} = \theta_{\ell,\kappa-1} - \frac{\alpha_\kappa}{\sqrt{\hat{v}_{\ell,\kappa}} + \epsilon} \hat{m}_{\ell,\kappa} \qquad (12)$$

where $\ell = 1, 2, \ldots, L$.

By using the Adam optimizer delineated in this section, we perform the parameter optimization of our DCNN-based classifiers (see Section V), where our objective function is the minimization of the empirical risk $R_J(\mathcal{C}_f)$ as defined in (8).

III. EXPERIMENTAL SETUP AND THE PROPOSED SOLUTION

In the following, we develop a more pragmatic and complex HAR system suitable for detecting human activities with motion in multiple directions. To this end, we utilize the multi-perspective 2×2 distributed MIMO radar configuration [53] (see Fig. 1) to eventually realize a direction-independent HAR system. The human activities were monitored by using the 2×2 MIMO radar configuration shown in Fig. 1. This configuration is also used for comparison with conventional SISO radar-based HAR systems, and to find out whether the multiperspective MIMO radar configuration can mitigate their limitations. We deployed a software-defined radar system known as Ancortek SDR-KIT 2400T2R4, which is an FMCW mmwave MIMO radar system, and used its transmitter-receiver antennas in a 2×2 configuration. The operating parameters of the Ancortek radar system are delineated in Table I.

For the proposed 2×2 MIMO radar-based HAR system, we arrange two radar subsystems, denoted by Radar$_1$ and Radar$_2$, where each radar subsystem has a collocated transmitter and a receiver antenna in a monostatic configuration. Radar$_1$ and Radar$_2$ are distributed in an indoor setting such that the 2×2 MIMO radar system renders a multiperspective illumination of a target as shown in Fig. 1, thereby having the potential to overcome the limitations that are posed by the monostatic SISO or monostatic MIMO radar systems in the context of HAR. We operate Radar$_1$ and Radar$_2$ in different time slots according to the TDMA scheme, where both radar subsystems have identical but independent radar
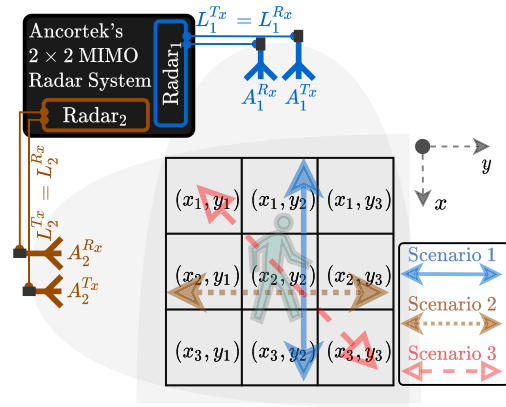

Fig. 1. Measurement setup of the proposed 2×2 MIMO radar-based HAR system consisting of Radar$_1$ and Radar$_2$, where Scenarios 1–3 characterize human activities in different directions.

TABLE I
2×2 MIMO RADAR SYSTEM PARAMETERS

Description	Symbols	Values
Carrier frequency	f_c	24.125 GHz
Radar bandwidth	BW	250 MHz
Chirp duration	T_{sw}	500 μs
Pulse repetition frequency	PRF	1 kHz
RF cable lengths	$(L_1^{Tx}, L_1^{Rx}, L_2^{Tx}, L_2^{Rx})$	$(0.3, 0.3, 7, 7)$ m

signal preprocessing chains (see Section II). The radar signal preprocessing chains process the raw in-phase and quadrature (IQ) data recorded by the Ancortek MIMO radar system. For a human activity, the radar signal preprocessing block of Radar$_i$ generates the TV radial velocity distribution $p_{ii}(v, t)$ by using (7) for $i \in \{1, 2\}$.

We consider five different types of human activities, which are as follows: falling on a mattress on the floor, walking, standing up from a chair, sitting down on a chair, and picking up an object from the floor. For these activities, the measured TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ are shown in Figs. 2 and 3, where the Scenarios 1, 2, and 3 denote the directions of human activities according to Fig. 1. In Scenario 1 (Scenario 2), the human motion is parallel to the boresight of Radar$_1$ (Radar$_2$), whereas in Scenario 3, the human movement is roughly at $45°$ to the boresights of both radar subsystems, as depicted in Fig. 1.

Radar$_1$ and Radar$_2$ complement each other such that when the activity direction changes from the x-axis to the y-axis of Fig. 1, the activity signature slowly vanishes from the radial velocity distribution $p_{11}(v, t)$ of Radar$_1$ and starts appearing in the radial velocity distribution $p_{22}(v, t)$ of Radar$_2$. For "Fall" activities performed in different directions, the measured radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ in the three scenarios vary significantly, as shown in Fig. 2. We can see from Figs. 2 and 3 that Radar$_1$ and Radar$_2$ are unable to acquire optimal human activity signatures in Scenarios 2 and 1, respectively. The suboptimal human activity signatures contribute toward the poor classification performance of a SISO radar-based direction-independent HAR system (see Section V-B). Therefore, analogous to a monostatic SISO or monostatic MIMO radar case, a single radial velocity distribution either
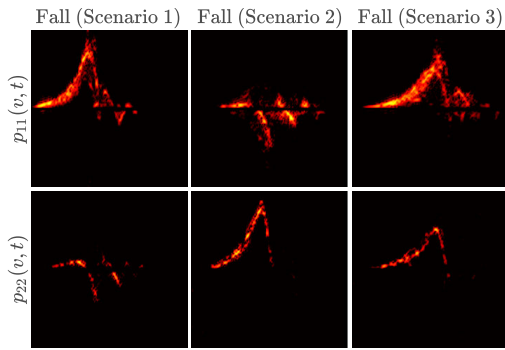
Fig. 2. Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Fall" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over 2–4 s.
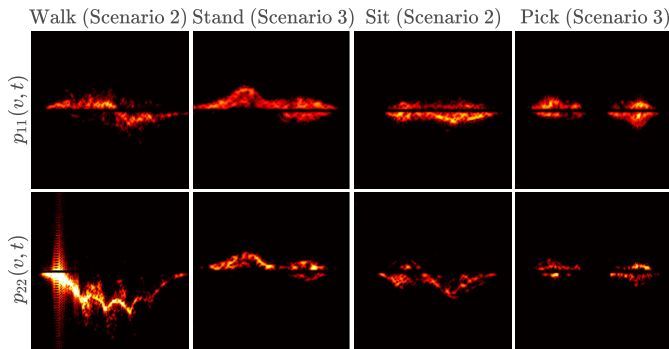

Fig. 3. Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of different human activities, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over 3–5 s.

from Radar$_1$ or Radar$_2$ cannot completely portray a human activity and would not be sufficient for the realization of a direction-independent HAR system. Additionally, Fig. 3 shows how the TV radial velocity distribution $p_{ii}(v, t)$ changes with the type of human activity. This figure demonstrates that the human activity signature or TV radial velocity distribution $p_{ii}(v, t)$ depends on the type as well as the direction of the human activity. To see the radar signatures corresponding to different multidirectional human activities, please refer to Figs. 12–15 in Appendix.

As our 2×2 distributed MIMO radar-based HAR system generates two distinct activity signatures from two different aspect angles for human activity, we must fuse or merge the information from the two activity signatures in order to accurately classify the human activity regardless of the direction of motion. In this research, we have implemented a fusion technique at the feature level. For this purpose, for each radar subsystem, Radar$_1$ and Radar$_2$, the features are extracted independently and automatically by several convolutional layers from the radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$, respectively. The extracted features from the radar subsystems are then merged by the concatenation layer (see Section V-C).

In Section V, we show how the classification performance of the monostatic radar-based HAR system deteriorates if the human activities take place in the 2-D xy plane, which is

depicted by the 3×3 grid in Fig. 1. We also explain the design of the proposed 2×2 distributed MIMO radar-based HAR system and show how it overcomes the above constraints on the direction of human activity motion. Compared to the SISO radar-based direction-independent HAR system, we see that the proposed 2×2 distributed MIMO radar-based direction-independent HAR system significantly ameliorates the classification accuracy.

IV. DATA COLLECTION

A comprehensive measurement campaign was carried out in an indoor environment consisting of fixed objects, such as chairs, tables, cabinets, computers, and other electronic items. The five types of activities were performed by six different persons, one of them was a female candidate. The human activities were carried out in several directions, with different speeds, and in different locations. For instance, the falling activities were performed in six different directions as depicted by the scenario markers in Fig. 1. Specifically, the falling activities were executed in the following directions: from (x_3, y_2) to (x_1, y_2), from (x_1, y_2) to (x_3, y_2), from (x_2, y_3) to (x_2, y_1), from (x_2, y_1) to (x_2, y_3), from (x_3, y_3) to (x_1, y_1), and from (x_1, y_1) to (x_3, y_3). The walking activities were performed and recorded in a similar fashion. The other human activities—standing up, sitting down, and picking up an object—were performed accordingly.

In this article, the term HAR$^{(2\text{-D})}$ is coined to represent the dataset recorded by the 2×2 MIMO radar system, where the superscript "(2-D)" refers to the human movement in the 2-D horizontal xy plane in Fig. 1. Therefore, for the direction-independent HAR task, we define FEN$^{(2\text{-D})}$, SISO$^{(2\text{-D})}$, and MIMO$^{(2\text{-D})}$ as a feature extraction network, a SISO radar-based HAR classifier, and a MIMO radar-based HAR classifier, respectively. On the other hand, to denote the human movement along the 1-D x-axis of the 3×3 grid in Fig. 1, we use the superscript "(1-D)." Thus, for the conventional 1-D HAR task, where the human movement is restricted to Scenario 1 in Fig. 1, we define HAR$^{(1\text{-D})}$, FEN$^{(1\text{-D})}$, and SISO$^{(1\text{-D})}$ as a dataset recorded by Radar$_1$, a feature extraction network, and a conventional SISO radar-based HAR classifier, respectively.

We need the HAR$^{(2\text{-D})}$ dataset to realize the SISO$^{(2\text{-D})}$ and MIMO$^{(2\text{-D})}$ HAR systems, whereas the HAR$^{(1\text{-D})}$ dataset is required for the SISO$^{(1\text{-D})}$ HAR system. The details of the HAR$^{(2\text{-D})}$ dataset related to the measurement campaign based on the proposed 2×2 MIMO radar framework are shown in Table II. As entered in Table II, we recorded a total of 1364 activities. For each activity, we generated the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ corresponding to the radar subsystems Radar$_1$ and Radar$_2$, respectively. On the other hand, Table III shows the HAR$^{(1\text{-D})}$ dataset, which is a proper subset of the HAR$^{(2\text{-D})}$ dataset, i.e., HAR$^{(1\text{-D})} \subset$ HAR$^{(2\text{-D})}$. Note that the HAR$^{(1\text{-D})}$ dataset only contains those activities of the HAR$^{(2\text{-D})}$ dataset that were performed parallel to the boresight of Radar$_1$. To implement a conventional monostatic radar-based HAR system (see Section V-A), we use only the TV radial velocity distributions $p_{11}(v, t)$ corresponding to the recorded human activities of the HAR$^{(1\text{-D})}$ dataset.

TABLE II
HAR$^{(2-D)}$ DATASET RECORDED BY A 2 × 2 MIMO RADAR SYSTEM

Type of human activity	Person #						Total trials
	1	2	3	4	5	6	
Falling	60	59	18	18	—	—	155
Walking	80	80	39	39	23	24	285
Standing up	102	100	27	27	27	27	310
Sitting down	105	102	27	23	22	25	304
Picking an object	101	103	25	27	27	27	310
Total trials	448	444	136	134	99	103	**1364**

TABLE III
HAR$^{(1-D)}$ DATA SUBSET RECORDED BY RADAR$_1$, WHERE THE
DIRECTION OF MOTION OF THE HUMAN ACTIVITIES
IS RESTRICTED TO MERELY SCENARIO 1

Type of human activity	Person #						Total trials
	1	2	3	4	5	6	
Falling	20	19	6	6	—	—	51
Walking	20	20	10	9	6	6	71
Standing up	35	32	9	9	9	9	103
Sitting down	35	32	9	7	8	8	99
Picking an object	34	33	9	9	9	9	103
Total trials	144	136	43	40	32	32	**427**

Each human activity trial was recorded for 10 s. The persons were told to maintain the initial and the final poses before and after performing the activity. Though each activity trial was recorded for 10 s, the actual duration of the activity was only 2–5 s, depending on the type of the activity and the speed at which the activity was carried out. We applied the active segment detection (ASD) [61] approach to the high-pass filtered in-phase component of the raw activity data to automatically extract an active segment, i.e., the section of the raw activity data corresponding to the actual duration of the activity. The ASD marks the start and end points of the activity by monitoring the variance of the filtered in-phase component of the raw activity data. The identified markers are used to extract active segments from the raw IQ activity data of Radar$_1$ and Radar$_2$. Thereafter, we applied radar signal processing techniques (see Section V-C) to compute the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ as given in (7).

To demonstrate the utility and effectiveness of our proposed multiperspective distributed MIMO radar approach, we develop three different types of classifiers or HAR systems. First, we develop a SISO$^{(1-D)}$ HAR system underlying a monostatic SISO radar configuration (see Section V-A). As conventional monostatic radar-based HAR systems only consider human activities performed along the radar boresight, SISO$^{(1-D)}$ uses the HAR$^{(1-D)}$ dataset for training and testing purposes. Second, to highlight how the classification performance of a HAR system deteriorates by the introduction of different movement directions, we developed a SISO radar-based direction-independent HAR system denoted as SISO$^{(2-D)}$ (see Section V-B). Unlike SISO$^{(1-D)}$, the SISO$^{(2-D)}$ HAR system makes use of the HAR$^{(2-D)}$ dataset for training and testing purposes because SISO$^{(2-D)}$ is designed to classify human activities in multiple directions of motion. Lastly, to significantly improve the classification performance of the SISO$^{(2-D)}$ HAR system, we also developed a 2 × 2 distributed MIMO radar-based direction-independent HAR system denoted as MIMO$^{(2-D)}$ (see Section V-C). Analogous to the SISO$^{(2-D)}$ HAR system, the proposed MIMO$^{(2-D)}$ HAR system uses the HAR$^{(2-D)}$ dataset for training and testing purposes, because MIMO$^{(2-D)}$ also considers the classification of human activities in multiple directions.

In this work, the recorded data from Person 1 and 2 were divided into training and validation datasets and used for the training phase of the DCNN-based SISO$^{(1-D)}$, SISO$^{(2-D)}$, and MIMO$^{(2-D)}$ classifiers. Of this data, 80% was used to train the classifiers, and 20% was used for validation. The recorded data from the rest of the participants—Person 3, 4, 5, and 6—were reserved to test the trained classifiers or HAR systems. In Sections V-A–V-C, we elucidate the design and development of the SISO$^{(1-D)}$, SISO$^{(2-D)}$, and MIMO$^{(2-D)}$ HAR systems, respectively, along with their results and discussions.

V. SISO AND DISTRIBUTED MIMO RADAR-BASED HAR SYSTEMS

A. Conventional SISO Radar-Based HAR System

In this section, we describe the design of the SISO$^{(1-D)}$ HAR system, which is analogous to a conventional SISO radar-based HAR system. We show the classification performance of the SISO$^{(1-D)}$ HAR system while restricting the human motion parallel to the boresight of Radar$_1$. Thus, we consider the HAR$^{(1-D)}$ dataset in Table III for the SISO$^{(1-D)}$ HAR system. Recall that the HAR$^{(1-D)}$ dataset contains only the human activities that were carried out in front of Radar$_1$ in Scenario 1. For all recorded human activities listed in Table III, we generated the TV radial velocity distributions $p_{11}(v, t)$ using the data of Radar$_1$ and converted the preprocessed data to images of size 224 × 224 × 3. Each image representing a human activity is a color image (see Figs. 2 and 3) with 224 pixels in the horizontal and vertical dimensions, and the number 3 refers to the red, green and blue (RGB) color channels.

The images of the radial velocity distributions $p_{11}(v, t)$ are used as input feature maps for the feature extraction network FEN$^{(1-D)}$ as depicted in Fig. 4. We can see from Fig. 4, that the first, second, and third convolutional layers of FEN$^{(1-D)}$ contain 32, 48, and 64 filter channels, respectively. The dimension of each 2-D learnable filter or kernel, also commonly known as kernel dimension k_d, is 6 × 6 pixels. For each convolutional layer of the SISO$^{(1-D)}$ network, we set the stride parameter to 1 so that the kernels are moved or strode by one pixel at a time. To avoid the problem of overfitting, we used $L2$ regularization [62] to penalize and eventually eliminate the spike-like weight vectors. The problems of slow convergence and vanishing gradients were mitigated by using the ReLU activation function on the convolutional layers [58]. Furthermore, each convolutional layer in Fig. 4 is followed by a max-pool layer and a dropout layer. The max-pool layer is of the order 2 × 2, which downsamples the output of the convolutional layer by a factor of 2. Each max-pool layer is followed by a dropout layer with a dropout rate of 15%. Finally, all the features that are generated by FEN$^{(1-D)}$ are flattened before feeding them to the fully connected layers.
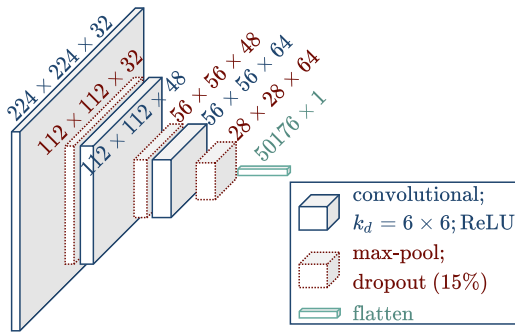
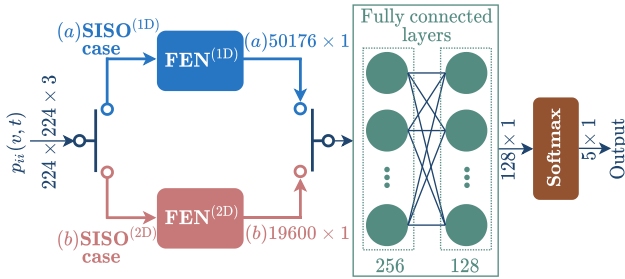Fig. 4. Feature extraction network FEN$^{(1\text{-}D)}$ designed for the SISO$^{(1\text{-}D)}$ HAR system.

Fig. 5. Architecture of the DCNN classifier for the monostatic SISO radar-based HAR systems, where (a) SISO$^{(1\text{-}D)}$ uses FEN$^{(1\text{-}D)}$ that outputs a feature vector of dimension $50\,176 \times 1$ and (b) SISO$^{(2\text{-}D)}$ uses FEN$^{(2\text{-}D)}$ that outputs a feature vector of dimension $19\,600 \times 1$.

A DCNN-based SISO$^{(1\text{-}D)}$ classifier is depicted in Fig. 5, where FEN$^{(1\text{-}D)}$ generates features from the input feature maps or, equivalently, the TV radial velocity distribution $p_{11}(v, t)$. Then, the extracted features undergo two fully connected layers of the order 256×1 and 128×1. As we are classifying five different types of human activities, the second-to-last fully connected layer is followed by an output layer of order 5×1 with the softmax activation function that converts the logits computed by the network into probabilities. To train the SISO$^{(1\text{-}D)}$ classifier, the HAR$^{(1\text{-}D)}$ dataset (see Table III) is divided into training, validation and testing datasets. The training and validation data account for 65.6% of the total data and belong to Persons 1 and 2, while the test data account for 34.4% of the HAR$^{(1\text{-}D)}$ dataset belonging to Persons 3, 4, 5, and 6.

In the training phase of the SISO$^{(1\text{-}D)}$ HAR system, we used the Adam optimizer to minimize the empirical risk $R_J(\mathcal{C}_f)$ in (8) corresponding to the categorical cross-entropy loss function J_{CCE}. Thus, the weights and biases of the DCNN-based SISO$^{(1\text{-}D)}$ classifier were optimized by using the Adam optimizer and the examples from the HAR$^{(1\text{-}D)}$ dataset. The default values of the decay factors or forgetting factors in (10) and (11) are equal to $\beta_1 = 0.9$ and $\beta_2 = 0.999$, respectively. In order to prevent division by 0 in (12), the value of ϵ was set to be 10^{-8}. A batch size of 32 was adopted in the training phase of the SISO$^{(1\text{-}D)}$ classifier. Note that the parameter optimization or training of the three classifiers—SISO$^{(1\text{-}D)}$, SISO$^{(2\text{-}D)}$, and MIMO$^{(2\text{-}D)}$—was performed in the same way with the same values for the network hyperparameters. For all three classifiers, the training history is summarized by the training loss, training accuracy, validation loss, and validation

Fig. 6. Confusion matrix of the results obtained by the SISO$^{(1\text{-}D)}$ HAR system. The first five entries of the last row and last column show the precision and recall, respectively, whereas the last entry highlighted in dark gray shows the overall accuracy.

accuracy curves in Fig. 10. During the training phase, which spans 100 epochs, there is no evidence of overfitting of the SISO$^{(1\text{-}D)}$ classifier (see Fig. 10).

We use a confusion matrix shown in Fig. 6 to summarize and quantitatively assess the overall performance of the trained DCNN-based SISO$^{(1\text{-}D)}$ classifier. The human activity classification performance of SISO$^{(1\text{-}D)}$ was evaluated using test-examples from the HAR$^{(1\text{-}D)}$ dataset. On the y-axis of the confusion matrix, we have the true class of an activity, and the x-axis shows the predicted class of an activity. Thus, for the first five rows and columns of the confusion matrix in Fig. 6, the diagonal entries show the number of correctly classified human activities, while the nondiagonal entries show the number of misclassified human activities. For example, the first column of the third row shows that a "Stand" activity has been incorrectly predicted or misclassified as a "Fall" activity. Moreover, the first five entries of the last row and last column of the confusion matrix show the precision and recall [63], respectively. Thus, we can see from Fig. 6 that the walking activity has a 100% recall and a precision of 96.88%. Most importantly, the overall accuracy of the SISO$^{(1\text{-}D)}$ classifier is 97.28%, which is indicated by the white color of the sixth entry in the last row and last column of the confusion matrix. It should be noted that using a complex network architecture (FEN$^{(2\text{-}D)}$) for a smaller dataset (HAR$^{(1\text{-}D)}$) can lead to overfitting and reduced generalizability. When we conducted experiments by changing the structure of FEN$^{(1\text{-}D)}$ to FEN$^{(2\text{-}D)}$ for the SISO$^{(1\text{-}D)}$ HAR system, as expected, we observed a small decline in the accuracy of the SISO$^{(1\text{-}D)}$ classifier, which dropped to 96.60% from 97.28%.

In this section, we looked into a conventional SISO radar-based HAR system denoted as SISO$^{(1\text{-}D)}$ that demonstrated a good classification performance (see Fig. 6). The classification performance of the SISO$^{(1\text{-}D)}$ classifier is comparable to state-of-the-art HAR systems. Analogous to the SISO$^{(1\text{-}D)}$ classifier or HAR system, most modern HAR systems that are based on either radar or Wi-Fi data are able to classify basic human activities with classification accuracies above 90% [24], [43], [64]. However, in these conventional monostatic

radar-based HAR systems, the human subjects' movements are limited to Scenario 1. In Section V-B, we extend the HAR problem by considering human motion in the horizontal xy plane, and investigate how this affects the classification performance of a conventional SISO radar-based direction-independent HAR system.

B. Direction-Independent SISO Radar-Based HAR System

To provide a comprehensive analysis and ensure a fair comparison, we include the SISO$^{(2-D)}$ approach in this section, which is a direction-independent monostatic SISO radar-based HAR system. This inclusion allows us to highlight the limitations of the SISO$^{(2-D)}$ HAR system and emphasize the effectiveness of the proposed MIMO$^{(2-D)}$ HAR system in addressing diverse directions of human activities. By comparing their performance using the HAR$^{(2-D)}$ dataset, we aim to demonstrate the significance of the proposed direction-independent HAR framework. Hence, we use the HAR$^{(2-D)}$ dataset as shown in Table II to realize the SISO$^{(2-D)}$ HAR system. For all the recorded human activities listed in Table II, we generated the TV radial velocity distributions $p_{ii}(v, t)$ by using Radar$_i$ data, where i may be chosen as either 1 or 2. For brevity, we report only the results of the SISO$^{(2-D)}$ HAR system trained and tested with the data of Radar$_1$. The TV radial velocity distributions $p_{11}(v, t)$ representing human activity fingerprints were converted into images of the order $224 \times 224 \times 3$ (see Figs. 2 and 3), which were used as input feature maps to the feature extraction network FEN$^{(2-D)}$ as depicted in Fig. 7.

The neural network architecture of the SISO$^{(2-D)}$ classifier is similar to the SISO$^{(1-D)}$ classifier except for a few modifications. For instance, the DCNN-based SISO$^{(2-D)}$ classifier uses FEN$^{(2-D)}$ instead of FEN$^{(1-D)}$ to extract features from the input feature maps or the TV radial velocity distribution $p_{ii}(v, t)$ as shown in Fig. 5. Compared with FEN$^{(1-D)}$ in Fig. 4, we see that FEN$^{(2-D)}$ in Fig. 7 has an additional convolutional layer, and each convolutional layer has a larger number of filters, i.e., 40, 60, 80, and 100. Consequently, the SISO$^{(2-D)}$ HAR system has a greater network complexity and capacity compared to the SISO$^{(1-D)}$ HAR system. Note that we needed a more complex DCNN classifier with higher network capacity because: 1) SISO$^{(2-D)}$ uses a larger HAR$^{(2-D)}$ dataset containing 1364 human activity fingerprints instead of 427 and 2) because SISO$^{(2-D)}$ aims to classify human activities in different directions, taking into account more diverse, complex, and sometimes suboptimal human activity signatures.

Moreover, the kernel dimension k_d of each 2-D learnable filter in FEN$^{(2-D)}$ is 5×5 as shown in Fig. 7. The rest of the specifications of the SISO$^{(2-D)}$ and SISO$^{(1-D)}$ classifiers are similar in terms of the max-pool layers, dropout layers, stride, batch size, and activation function. Analogous to SISO$^{(1-D)}$, SISO$^{(2-D)}$ uses $L2$ regularization to penalize and eliminate the peaky weight vectors to avoid the overfitting problem. Like SISO$^{(1-D)}$, SISO$^{(2-D)}$ uses the Adam optimizer to minimize the empirical risk $R_J(\mathcal{C}_f)$ in (8) corresponding to the categorical cross-entropy loss function J_{CCE}. In order to train the SISO$^{(2-D)}$
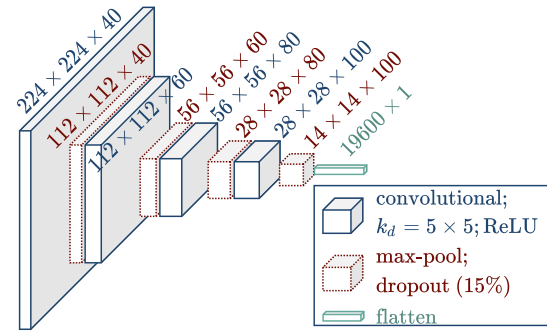

Fig. 7. Feature extraction network FEN$^{(2-D)}$ designed for SISO$^{(2-D)}$ and MIMO$^{(2-D)}$ HAR systems.

classifier, the HAR$^{(2-D)}$ dataset is split into training, validation, and testing datasets. The training and validation data is 65.4% of the total data and belongs to Person 1 and 2, whereas the testing data is 34.6% of the HAR$^{(2-D)}$ dataset belonging to Person 3, 4, 5, and 6. As mentioned in Section V-A, the training history is summarized by the training loss, training accuracy, validation loss, and validation accuracy curves shown in Fig. 10 for all three classifiers. Note that for the SISO$^{(2-D)}$ classifier, there is no evidence of overfitting during the training phase that spans over 100 epochs, as shown in Fig. 10.

The classification performance of the SISO$^{(2-D)}$ direction-independent HAR system was evaluated using the test-examples from the HAR$^{(2-D)}$ dataset. Recall that the SISO$^{(2-D)}$ HAR system is realized by using the data of Radar$_1$. To summarize and quantitatively assess the classification performance of the SISO$^{(2-D)}$ HAR system, we present a confusion matrix in Fig. 8. The predicted and actual class of a human activity is shown on the x-axis and y-axis of the confusion matrix, respectively. The confusion matrix in Fig. 8 shows that the overall classification performance of the SISO$^{(2-D)}$ HAR system has dropped significantly to only 88.98%. On a partially unrelated note and without going into too much detail, we would also like to mention that Radar$_2$ provides relatively poor data quality due to the cross-channel interference problem [65]. Solving the cross-channel interference problem requires the deployment of longer RF cables (see Table I), which cause a higher attenuation of the received signal. For this reason, a SISO$^{(2-D)}$ direction-independent HAR system realized by using only the data of Radar$_2$ provided an overall classification accuracy of just 83.05%.

Looking at the nondiagonal entries of the confusion matrix in Fig. 8, we see numerous misclassified human activities, e.g., the "Pick" activity was misclassified 15 times as the "Stand" activity by the SISO$^{(2-D)}$ HAR system. Therefore, the worst precision of the system is 76.34% corresponding to the "Stand" activity, and the worst recall is observed as 80.19% for the "Pick" activity. Interestingly, the precision and recall are 100% for the "Fall" activity, which implies that the SISO$^{(2-D)}$ HAR system learned to classify the human falling activity in all directions. Unfortunately, this is not true for the other four types of human activity, which have diverse and relatively complex radial velocity distributions that vary in different directions (see Figs. 2 and 3).

	Fall	Walk	Stand	Sit	Pick	
Fall	36	0	0	0	0	100% 0%
Walk	0	116	9	0	0	92.8% 7.2%
Stand	0	3	100	3	2	92.59% 7.41%
Sit	0	4	7	83	3	85.57% 14.43%
Pick	0	0	15	6	85	80.19% 19.81%
	100% 0%	94.31% 5.69%	76.34% 23.66%	90.22% 9.78%	94.44% 5.56%	88.98% 11.02%
	Fall	Walk	Stand	Sit	Pick	

True Class (vertical axis); Predicted Class (horizontal axis)

Fig. 8. Confusion matrix of the results obtained by the SISO$^{(2\text{-D})}$ HAR system, where SISO$^{(2\text{-D})}$ was trained and tested by using Radar$_1$ data. The first five entries of the last row and last column show the precision and recall, respectively, whereas the last entry shows the overall accuracy.

In this section, a direction-independent SISO radar-based HAR system (SISO$^{(2\text{-D})}$) was investigated, which showed significant degradation in its classification performance for human motion in different directions. For the simpler case of human motion, or when the human motion was restricted to Scenario 1 in Fig. 1, the overall classification accuracy of the SISO$^{(1\text{-D})}$ HAR system was 97.28%. However, when we complicated the human motion by considering the different directions of motion, the classification accuracy dropped to 88.98% for the SISO$^{(2\text{-D})}$ HAR system. The deterioration of the classification performance manifested by the SISO$^{(2\text{-D})}$ HAR system comes from the physical limitations of monostatic SISO radar systems. These physical limitations of monostatic radar systems can be overcome by the 2×2 distributed MIMO radar configuration of Fig. 1 to eventually realize a direction-independent MIMO$^{(2\text{-D})}$ HAR system. In Section V-C, we will see how the MIMO$^{(2\text{-D})}$ HAR system ameliorates the shortcomings of the SISO$^{(1\text{-D})}$ and SISO$^{(2\text{-D})}$ HAR systems altogether.

C. 2×2 MIMO Radar-Based Direction-Independent HAR System

We now elucidate the design of our proposed 2×2 distributed MIMO radar-based direction-independent HAR system denoted as MIMO$^{(2\text{-D})}$. Considering the different directions of human activities in the horizontal xy plane in Fig. 1, we use the HAR$^{(2\text{-D})}$ dataset (see Table II) to eventually realize the MIMO$^{(2\text{-D})}$ HAR system. In this section, we demonstrate that unlike the SISO$^{(2\text{-D})}$ HAR system, our proposed MIMO$^{(2\text{-D})}$ HAR system is able to recognize the human activities with a very good classification performance for the HAR$^{(2\text{-D})}$ dataset. For all the recorded human activities listed in Table II, we computed the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ by using Radar$_1$ and Radar$_2$ data, respectively. The TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ were converted separately into images of the order $224 \times 224 \times 3$ (see Figs. 2 and 3), which served as input feature maps to the feature extraction network FEN$^{(2\text{-D})}$ as depicted in Fig. 7.

Although the neural network architecture of the MIMO$^{(2\text{-D})}$ and SISO$^{(2\text{-D})}$ HAR systems are quite different in Fig. 9 and Fig. 5, respectively, the building blocks, hyperparameter values, and training processes of the two networks are very similar. For instance, the MIMO$^{(2\text{-D})}$ and SISO$^{(2\text{-D})}$ HAR systems use the same specifications related to kernel dimension k_d, max-pool layers, dropout layers, stride, batch size, activation function, regularizer, and Adam optimizer (refer to Section V-B for more details). Moreover, the same feature extraction network FEN$^{(2\text{-D})}$ in Fig. 7 has been adopted for the MIMO$^{(2\text{-D})}$ and SISO$^{(2\text{-D})}$ HAR systems. However, unlike the SISO$^{(2\text{-D})}$ HAR system, the MIMO$^{(2\text{-D})}$ HAR system uses two identical feature extraction blocks as depicted in Fig. 9 for the TV radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$. The two FEN$^{(2\text{-D})}$ blocks of MIMO$^{(2\text{-D})}$ HAR system extract unique features automatically and independently of the two radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$. In Fig. 9, we can see that these features are then merged using a concatenation layer, which is followed by MLP and softmax layers to eventually classify the human activities.

Analogous to the SISO$^{(2\text{-D})}$ HAR system, the MIMO$^{(2\text{-D})}$ HAR system also uses the HAR$^{(2\text{-D})}$ dataset for the training and testing purposes. However, for the MIMO$^{(2\text{-D})}$ HAR system, the main difference is that the activity fingerprints from both radar subsystems shown in Fig. 1 are simultaneously utilized to classify the human activities. In other words, for the classification of human activity, two distinct multiperspective radial velocity distributions $p_{11}(v, t)$ and $p_{22}(v, t)$ produced by Radar$_1$ and Radar$_2$, respectively, are processed at once by the MIMO$^{(2\text{-D})}$ HAR system. Therefore, in the MIMO$^{(2\text{-D})}$ HAR system, we utilized 2728 images or equivalently 1364 pairs of images corresponding to 1364 human activities of the HAR$^{(2\text{-D})}$ dataset. The HAR$^{(2\text{-D})}$ dataset was split into training, validation, and testing datasets, where the training and validation data was 65.4% of the total data belonging to Person 1 and 2, and the testing data was 34.6% of the total data belonging to Person 3, 4, 5, and 6. Recall that the training history is summarized by the training loss, training accuracy, validation loss, and validation accuracy curves as depicted in Fig. 10 for all three classifiers or HAR systems. Note that this figure does not reveal any signs of overfitting during the training phase of the MIMO$^{(2\text{-D})}$ HAR system.

In Fig. 11, we present a confusion matrix to quantitatively assess the overall classification performance of the MIMO$^{(2\text{-D})}$ direction-independent HAR system. The human activity classification performance of the MIMO$^{(2\text{-D})}$ HAR system was evaluated over the test examples from the HAR$^{(2\text{-D})}$ dataset. In the test examples, the number of falling activities is comparatively low because it is difficult to carry out a real-life "Fall" activity. Nevertheless, the train–test split ratio is roughly $77 : 23$ for the "Fall" activity. In the confusion matrix in Fig. 11, the overall classification performance of the MIMO$^{(2\text{-D})}$ direction-independent HAR system comes out to be 98.52%, which is a significant improvement over the classification accuracy of 88.98% achieved by the SISO$^{(2\text{-D})}$ direction-independent HAR system. Looking at the nondiagonal entries of the confusion matrix in Fig. 11, we see only seven misclassified human activities. We can observe that the
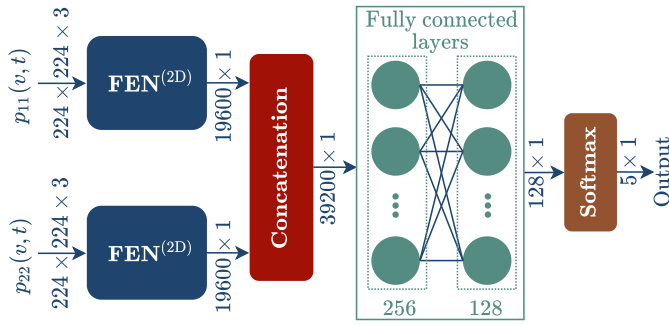
Fig. 9. Architecture of the proposed MIMO$^{(2\text{-}D)}$ HAR system with two independent FEN$^{(2\text{-}D)}$ blocks to generate feature vectors that are fused by the concatenation layer for subsequent classification.
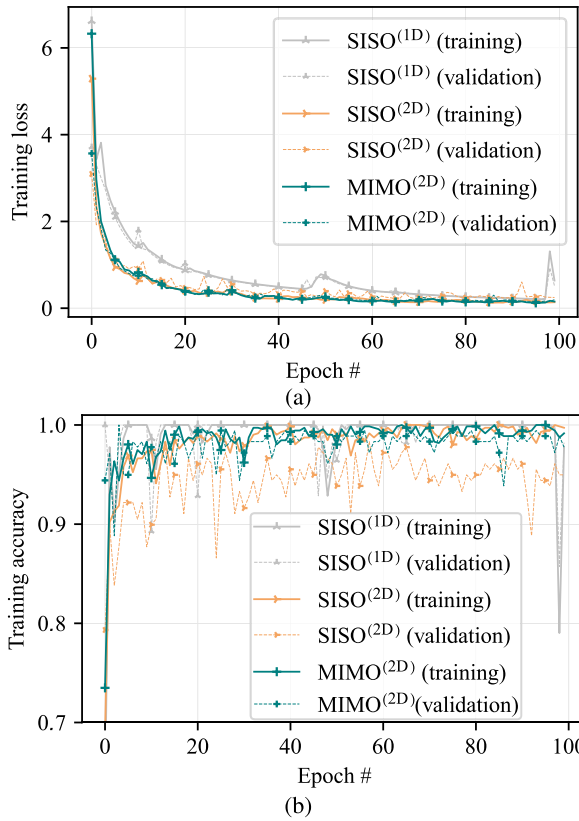

Fig. 10. Training history for the SISO$^{(1\text{-}D)}$, SISO$^{(2\text{-}D)}$, and MIMO$^{(2\text{-}D)}$ HAR systems. (a) Training losses. (b) Training accuracies over 100 epochs.

worst precision of the MIMO$^{(2\text{-}D)}$ HAR system is 95.54% corresponding to the "Stand" activity, and the worst recall is observed as 94.34% for the "Pick" activity. Note that the increase in the classification performance is basically due to the multiperspective illumination of the environment by the proposed 2×2 distributed MIMO radar-based HAR system.

We addressed a HAR task in a complex situation, where we considered the human motion in the horizontal xy plane in Fig. 1. To mitigate the shortcomings of the SISO$^{(2\text{-}D)}$ HAR system in relation to the human activity direction, we illuminated the subject from different aspect angles by using the proposed 2×2 MIMO radar-based direction-independent HAR system denoted as MIMO$^{(2\text{-}D)}$, which demonstrated a remarkably good classification performance as summarized by

Fig. 11. Confusion matrix of the results obtained by the proposed MIMO$^{(2\text{-}D)}$ HAR system with an overall accuracy of 98.52%.

the confusion matrix in Fig. 11. As evident from the classification performance of the MIMO$^{(2\text{-}D)}$ HAR system, the physical limitations of the monostatic radar systems were successfully mitigated by the multiperspective 2×2 distributed MIMO radar configuration. Therefore, by addressing and rectifying the fundamental radar problem at the physical layer, we were able to design a radar-based HAR system that was capable of recognizing human activities independent of their directions with a classification accuracy close to 100%.

VI. CONCLUSION

In this article, we analyzed and resolved a crucial physical layer problem of state-of-the-art monostatic SISO, SIMO, and MIMO radar-based HAR systems, which primarily arises due to the target's aspect angle. Thus, a more pragmatic and more complex HAR problem has been elucidated in this research in the context of RF sensing, where we improve the activity recognition task by considering multiple directions of human activities. A novel HAR dataset (HAR$^{(2\text{-}D)}$) was recorded by using the proposed multiperspective 2×2 MIMO radar framework. We developed and analyzed three different HAR systems, denoted as SISO$^{(1\text{-}D)}$, SISO$^{(2\text{-}D)}$, and MIMO$^{(2\text{-}D)}$, by using our HAR$^{(2\text{-}D)}$ dataset and its sub-dataset HAR$^{(1\text{-}D)}$.

Analogous to most modern radar-based HAR systems, the SISO$^{(1\text{-}D)}$ HAR system was able to classify human activities with a classification accuracy of 97.28%. However, in this conventional monostatic radar-based HAR approach, the movement of the human subjects was restricted along the radar's boresight axis. By developing and analyzing the monostatic SISO$^{(2\text{-}D)}$ HAR system and considering the human activities taking place in the 2-D xy plane, we substantiated a significant deterioration in the classification performance from 97.28% to 88.98%. The deterioration of the classification performance manifested by the SISO$^{(2\text{-}D)}$ HAR system came from the inherent physical layer limitations of the monostatic SISO radar systems. To overcome these physical layer issues and drawbacks experienced by today's radar-based HAR systems, we utilized a multiperspective 2×2 distributed MIMO radar system to realize a direction-independent HAR system that was capable of recognizing human gross motor

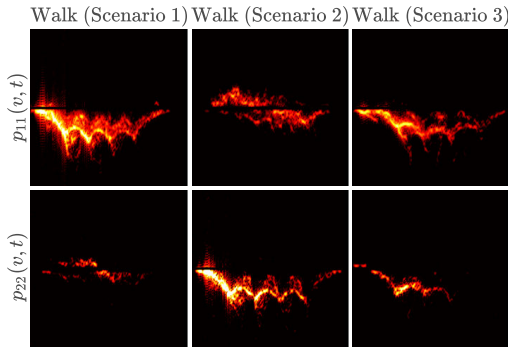Fig. 12. Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Walk" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over 3–5 s.
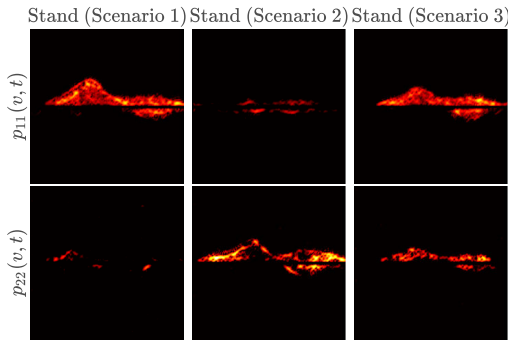

Fig. 13. Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Stand" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over 2–3 s.
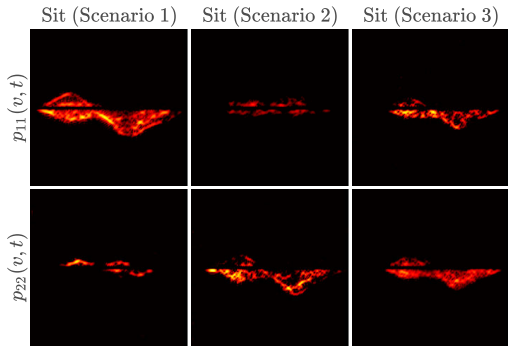

Fig. 14. Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Sit" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over 2–3 s.

activities regardless of the aspect angle or direction of motion. To eradicate the limitations that emerge due to the direction of motion, feature level fusion was adopted in the DCNN-based MIMO$^{(2\text{-}D)}$ classifier, which virtually combines the target's information from different aspect angles.

For the HAR$^{(2\text{-}D)}$ dataset, it was shown that the proposed multiperspective MIMO$^{(2\text{-}D)}$ HAR system significantly outperforms the monostatic SISO$^{(2\text{-}D)}$ HAR system. Compared with the SISO$^{(2\text{-}D)}$ HAR system, the proposed MIMO$^{(2\text{-}D)}$ HAR system significantly improved the classification accuracy from 88.98% to 98.52%. Therefore, the physical layer limitations of the monostatic SISO radar-based HAR systems were successfully mitigated by the proposed MIMO$^{(2\text{-}D)}$ HAR system.
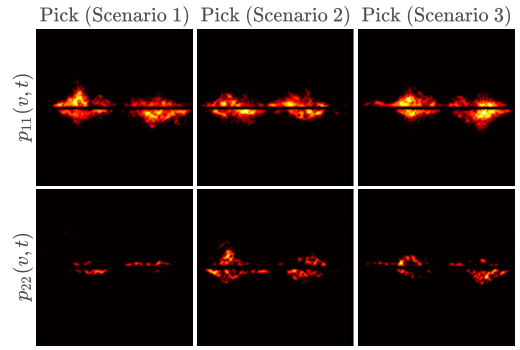

Fig. 15. Images containing the heatmap of the measured radial velocity distributions $p_{ii}(v,t)$ of the "Pick" activity in three different scenarios, where each image has the radial velocity v on the y-axis ranging $[-1.5, 1.5]$ m/s and time t on the x-axis spanning over 2–4 s.

The MIMO$^{(2\text{-}D)}$ HAR system presented in this article paves a way forward toward actualizing a more realistic and more advanced radar-based HAR system. To further enhance the classification performance, we plan to use the bistatic components of the 2×2 MIMO radar system, which are the TV radial velocity distributions $p_{12}(v, t)$ and $p_{21}(v, t)$. For more aspect angle coverage and a more complex HAR problem, we plan to extend the fundamental distributed 2×2 MIMO radar system to a larger MIMO antenna configuration.

APPENDIX
MULTIDIRECTIONAL HUMAN ACTIVITY SIGNATURES

In this appendix, we provide the measured radial velocity distributions $p_{ii}(v, t)$ of the four types of human activities: walking, standing up from a chair, sitting down on a chair, and picking up an object from the floor.

REFERENCES

[1] F. Gu, M. H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, "A survey on deep learning for human activity recognition," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–34, 2022.

[2] M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *Npj Digit. Med.*, vol. 4, no. 1, p. 148, Oct. 2021.

[3] Y. Guo et al., "Evolutionary dual-ensemble class imbalance learning for human activity recognition," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 4, pp. 728–739, Aug. 2022.

[4] Z. Ma, L. T. Yang, M. Lin, Q. Zhang, and C. Dai, "Weighted support tensor machines for human activity recognition with smartphone sensors," *IEEE Trans. Ind. Informat.*, early access, Feb. 23, 2021, doi: 10.1109/TII.2021.3061559.

[5] F. J. Rodriguez Lera, F. M. Rico, A. M. G. Higueras, and V. M. Olivera, "A context-awareness model for activity recognition in robot-assisted scenarios," *Exp. Syst.*, vol. 37, no. 2, Apr. 2020, Art. no. e12481.

[6] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, Jun. 2021.

[7] K. Host and M. Ivasic-Kos, "An overview of human action recognition in sports based on computer vision," *Heliyon*, vol. 8, no. 6, Jun. 2022, Art. no. e09633.

[8] T. Tuncer, F. Ertam, S. Dogan, and A. Subasi, "An automated daily sports activities and gender recognition method based on novel multikernel local diamond pattern using sensor signals," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9441–9448, Dec. 2020.

[9] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Elect. Eng. Technol.*, vol. 14, pp. 2567–2573, Sep. 2019.

[10] M. Al-Khafajiy et al., "Remote health monitoring of elderly through wearable sensors," *Multimedia Tools Appl.*, vol. 78, no. 17, pp. 24681–24706, Sep. 2019.

[11] M. Rashmi, T. S. Ashwin, and R. M. R. Guddeti, "Surveillance video analysis for student action recognition and localization inside computer laboratories of a smart campus," *Multimedia Tools Appl.*, vol. 80, no. 2, pp. 2907–2929, Jan. 2021.

[12] F. Gurcan, N. E. Cagiltay, and K. Cagiltay, "Mapping human-computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years," *Int. J. Hum.-Comput. Interact.*, vol. 37, no. 3, pp. 267–280, Feb. 2021.

[13] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowl.-Based Syst.*, vol. 223, Jul. 2021, Art. no. 106970.

[14] J. Roche, V. De-Silva, J. Hook, M. Moencks, and A. Kondoz, "A multimodal data processing system for LiDAR-based human activity recognition," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10027–10040, Oct. 2022.

[15] W. Qi, N. Wang, H. Su, and A. Aliverti, "DCNN based human activity recognition framework with depth vision guiding," *Neurocomputing*, vol. 486, pp. 261–271, May 2022.

[16] M. A. Al-Qaness, A. Dahou, M. A. Elaziz, and A. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 144–152, Jan. 2023.

[17] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 2106–2116, Feb. 2023.

[18] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, "Triple cross-domain attention on human activity recognition using wearable sensors," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 5, pp. 1167–1176, Oct. 2022.

[19] W. Huang, L. Zhang, W. Gao, F. Min, and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[20] L. Zhang, W. Cui, B. Li, Z. Chen, M. Wu, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1765–1775, Mar. 2023.

[21] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 487–499, Jan. 2023.

[22] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5103112.

[23] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 1, pp. 286–293.

[24] M. Muaaz, A. Chelli, and M. Pätzold, "WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–7.

[25] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019.

[26] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[27] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2016, pp. 363–373.

[28] Z. Peng et al., "A portable FMCW interferometry radar with programmable low-IF architecture for localization, ISAR imaging, and vital sign tracking," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 4, pp. 1334–1344, Apr. 2017.

[29] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, Mar. 2022.

[30] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *Proc. 15th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2018, pp. 805–809.

[31] L. Pallotta, M. Cauli, C. Clemente, F. Fioranelli, G. Giunta, and A. Farina, "Classification of micro-Doppler radar hand-gesture signatures by means of Chebyshev moments," in *Proc. IEEE 8th Int. Workshop Metrol. Aerosp. (MetroAeroSpace)*, Jun. 2021, pp. 182–187.

[32] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina, "A novel algorithm for radar classification based on Doppler characteristics exploiting orthogonal pseudo-Zernike polynomials," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 417–430, Jan. 2015.

[33] M. B. Özcan, S. Z. Gürbüz, A. R. Persico, C. Clemente, and J. Soraghan, "Performance analysis of co-located and distributed MIMO radar for micro-Doppler classification," in *Proc. Eur. Radar Conf. (EuRAD)*, Oct. 2016, pp. 85–88.

[34] L. Pallotta, C. Clemente, A. De Maio, J. J. Soraghan, and A. Farina, "Pseudo-Zernike moments based radar micro-Doppler classification," in *Proc. IEEE Radar Conf.*, May 2014, pp. 0850–0854.

[35] B. Erol and M. Amin, "Effects of range spread and aspect angle on radar fall detection," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.

[36] C. Ding et al., "Continuous human motion recognition with a dynamic range-Doppler trajectory method based on FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6821–6831, Sep. 2019.

[37] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion signature," in *Proc. 5th Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth)*, May 2011, pp. 222–225.

[38] L. Liu, M. Popescu, K. C. Ho, M. Skubic, and M. Rantz, "Doppler radar sensor positioning in a fall detection system," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 256–259.

[39] P. Nallabolu, L. Zhang, H. Hong, and C. Li, "Human presence sensing and gesture recognition for smart home applications with moving and stationary clutter suppression using a 60-GHz digital beamforming FMCW radar," *IEEE Access*, vol. 9, pp. 72857–72866, 2021.

[40] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *Proc. 13th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2016, pp. 671–676.

[41] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *Proc. IEEE Radar Conf.*, 2015, pp. 1491–1496.

[42] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, Jun. 2022.

[43] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "RadHAR: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proc. Annu. Int. Conf. Mob. Comput. Netw.*, 2019, pp. 51–56.

[44] J. Pan, M. Sun, Y. Wang, and X. Zhang, "An enhanced spatial smoothing technique with ESPRIT algorithm for direction of arrival estimation in coherent scenarios," *IEEE Trans. Signal Process.*, vol. 68, pp. 3635–3643, 2020.

[45] Z. Wang, Z. Yang, S. Wu, H. Li, S. Tian, and X. Chen, "An improved multiple signal classification for nonuniform sampling in blade tip timing," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7941–7952, Oct. 2020.

[46] M. Jian, Z. Lu, and V. C. Chen, "Drone detection and tracking based on phase-interferometric Doppler radar," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2018, pp. 1146–1149.

[47] S. Kim, B.-S. Kim, Y. Jin, and J. Lee, "Extrapolation-RELAX estimator based on spectrum partitioning for DOA estimation of FMCW radar," *IEEE Access*, vol. 7, pp. 98771–98780, 2019.

[48] F. Fioranelli, M. Ritchie, and H. Griffiths, "Bistatic human micro-Doppler signatures for classification of indoor activities," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 610–615.

[49] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.

[50] B. Jokanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, Feb. 2018.

[51] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, Jan. 2019.

[52] M. Bekar, C. J. Baker, E. G. Hoare, and M. Gashinova, "Joint MIMO radar and communication system using a PSK-LFM waveform with TDM and CDM approaches," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6115–6124, Mar. 2021.

[53] S. Waqar, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Appl. Sci.*, vol. 12, no. 4, p. 1825, Feb. 2022.

[54] M. Muaaz, S. Waqar, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *Proc. 4th Int. Conf. Intell. Technol. Appl.*, 2022, pp. 259–272.

[55] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.

[56] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 8792–8802.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[58] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2684–2691.

[59] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Sep. 2014.

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, Dec. 2015, pp. 1–15.

[61] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: From authentication to imitation," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3209–3221, Nov. 2017.

[62] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2019, pp. 1–19.

[63] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.

[64] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-sense: A passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, nos. 3–4, pp. 163–175, Apr. 2022.

[65] S. Waqar and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, p. 7496, Nov. 2021.

Sahil Waqar received the bachelor's (Hons.) degree in electrical engineering from the University of Engineering and Technology (UET), Taxila, Pakistan, in 2010, and the master's (Hons.) degree in electrical engineering from the National University of Science and Technology (NUST), Islamabad, Pakistan, in 2013. His master's research was related to probabilistic data association filters.

From 2012 to 2020, he has worked on the design and development of digital communication systems, signal processors, radar systems, and embedded systems. Since October 2020, he has been working on the CareWell Project as a Ph.D. Research Fellow with the University of Agder, Grimstad, Norway. His current research interests include machine/deep learning, sensors, multiple-input–multiple-output (MIMO) radars, communications, signal processing, and embedded systems.

Mr. Waqar has received the Gold Medal Award for his master's degree.

Muhammad Muaaz received the M.Sc. degree in information and communication systems security from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2012, and the Ph.D. degree in computer science from Johannes Kepler University, Linz, Austria, in 2017.

From 2018 to 2020, he held postdoctoral fellowships at the University of Agder, Grimstad, Norway, and Johannes Kepler University. He is currently working as a Researcher with the University of Agder. His research interests include the design and development of secure intelligent systems, information and communication systems security, data science, machine (deep) learning, mobile computing, and physiological and behavioral biometrics.

Matthias Pätzold (Senior Member, IEEE) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Ruhr University Bochum, Bochum, Germany, in 1985 and 1989, respectively, and the Habilitation degree in communications engineering from the Hamburg University of Technology, Hamburg, Germany, in 1998.

From 1990 to 1992, he has worked with ANT Nachrichtentechnik GmbH, Backnang, Germany, where he was involved in digital satellite communications. From 1992 to 2001, he was with the Department of Digital Networks, Hamburg University of Technology. Since 2001, he has been a Full Professor of Mobile Communications with the University of Agder, Grimstad, Norway. He has authored several books and numerous technical articles.

Dr. Pätzold has been actively participating in numerous conferences as a TPC member and the TPC chair. His publications received 14 Best Paper Awards.