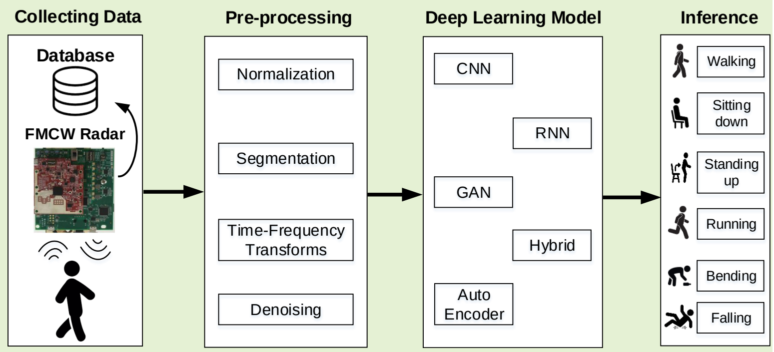
Deep Learning-based Human Activity Recognition with FMCW Radar: A Review

Van Ngoc Dang, *Student Member, IEEE*, Ngoc Chau Hoang, Minh Thuy Le, *Senior Member, IEEE*, Kien Nguyen, *Senior Member, IEEE*, and Quoc Cuong Nguyen, *Member, IEEE*



***Abstract*—Human activity recognition (HAR) has emerged as a critical research area with strong implications for healthcare, including elderly monitoring in assisted living and independent environments. Although several surveys have examined radar-based HAR, no comprehensive review has focused specifically on deep learning methods using frequency-modulated continuous-wave (FMCW) radar. To address this gap, we systematically analyzed 82 peer-reviewed publications spanning 2019–2025 from leading digital libraries. Our findings reveal a rapid growth in deep learning–enabled FMCW radar HAR, with major themes including activity classification, fall detection, and radar-based sensing for healthcare and IoT contexts. State-of-the-art models leverage convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders (AEs), and hybrid architectures to extract features from range–time, range–Doppler, micro-Doppler, range–angle, and point-cloud domains. Despite notable progress, open challenges remain in computational complexity, limited public datasets, inter-class similarity, environmental robustness, and the absence of standardized evaluation frameworks. This survey synthesizes current advances and identifies research directions, providing practical guidance for researchers and practitioners developing next-generation FMCW radar–based HAR systems.**

***Index Terms*—Human activity recognition, deep learning, frequency modulated continuous wave (FMCW) radar, PRISMA.**

I. Introduction

THE rapid aging of the global population [1] has intensified concerns about healthcare and safety, particularly for older adults who face elevated risks of falls. Falls remain a leading cause of injury, psychological trauma, and mortality among the elderly, underscoring the urgent need for monitoring systems capable of timely intervention in independent living or assisted-care settings. With the advancement of sensor technologies, human activity recognition (HAR) has emerged as a promising approach for monitoring daily activities and detecting emergencies [2]–[4]. HAR enables the automated recognition of human movements, supports early detection of adverse events such as falls, and reduces caregiver burden by providing continuous health monitoring. Beyond healthcare, HAR applications extend to smart homes [5], [6], security [7], automotive safety [8]–[10], industrial worker monitoring [11], [12], and human–computer interaction (HCI) [13], [14]. In smart homes, HAR informs environmental control and appliance automation. In security, it detects abnormal activities in restricted areas. In automotive systems, it identifies risky driver behaviors (e.g., drowsiness or phone use). Industrial deployments monitor worker activities for safety and performance, while in HCI, HAR enables natural interfaces through gesture and body-motion recognition.

Existing HAR systems have traditionally relied on cameras [15]–[17] and wearable sensors [18]–[20]]. Camera-based HAR provides rich visual data but suffers from lighting sensitivity, occlusions, and persistent privacy concerns. Wearable sensors offer real-time monitoring but face challenges of user discomfort and battery limitations. To overcome these drawbacks, frequency-modulated continuous-wave (FMCW) radar has gained traction as a privacy-preserving, lighting-invariant sensing modality. FMCW radar can operate reliably under varying environmental conditions, making it particularly attractive for healthcare and elderly monitoring applications.

Traditional machine learning (ML) methods such as decision trees [21], Hidden Markov model [22], support vector machines (SVM) [23], and random forests [24]have been widely applied to HAR. These algorithms allow effective feature engineering and classification on small datasets with

Van Ngoc Dang, Ngoc Chau Hoang, Minh Thuy Le, and Quoc Cuong Nguyen are with the Sensor Laboratory, School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi 100000, Vietnam (e-mail: ngoc.dv230048d@sis.hust.edu.vn; chau.hn222175m@sis.hust.edu.vn; thuy.leminh@hust.edu.vn; cuong.nguyenquoc@hust.edu.vn), (*Corresponding author: Quoc Cuong Nguyen*).

Kien Nguyen is with Chiba University, Japan. (e-mail: nguyen@chiba-u.jp).

modest computational demands. However, ML-based HAR depends heavily on manual feature extraction, limiting scalability and generalization to unseen data. In contrast, deep learning (DL) automatically learns hierarchical spatial–temporal features from raw data, improving generalization and robustness. DL architectures exploit large datasets and GPU acceleration to uncover complex motion patterns and support near real-time inference, positioning DL as the state-of-the-art approach for HAR across diverse environments.

Several surveys have reviewed HAR using radar or multimodal sensors [25]–[32]. Islam et al. [25] examined CNN-based HAR across smartphones, radar, and vision data.s. Hu et al. [26] focused on radar-based fall detection, while Tewari et al. [27] discussed radar sensing for non-invasive fall monitoring. Zhang et al. [28] summarized mmWave-based human sensing, covering localization, motion recognition, biometrics, and imaging. Ullmann et al. [29] addressed continuous HAR in real-world radar streams. Ahmed et al. [30] reviewed healthcare radar applications but lacked detailed analysis of deep learning. Nocera et al. [31] evaluated radar-based physiological monitoring using ML/DL, and Miazek et al. [32] considered multimodal behavior analysis including radar, LiDAR, and video. While these reviews provide important insights, they either emphasize specific applications (e.g., fall detection) or aggregate multiple modalities without a dedicated focus on deep learning–based HAR using FMCW radar. To date, no systematic survey has comprehensively examined how deep learning architectures exploit FMCW radar data for HAR. This review addresses this gap through a systematic analysis of 82 peer-reviewed studies published between 2018 and 2025 across major digital libraries (IEEE, ACM, Springer, Elsevier). Publications were selected using explicit inclusion and exclusion criteria, summarized insec2. Table I presents a comparison between this work and others. In this work, we investigate the following research questions (RQs):

- RQ1: What is the state of the art on deep learning-based HAR using FMCW radar?
- RQ2: What radar signal processing techniques are employed for HAR?
- RQ3: How have deep learning models evolved for FMCW radar HAR?
- RQ4:What public FMCW radar datasets exist to support reproducible research?

We have organized the rest of this review as follows. sec2 details the research methodology, including the inclusion and exclusion criteria applied to the selected publications. sec3 presents an analysis of the data to address the four proposed research questions. sec4 discusses the key challenges in the field and directions for future research. Finally, sec5 concludes the review.

II. Material and method

This study draws on three leading academic databases: ScienceDirect, IEEE Xplore, and Springer Nature. They have been chosen for their extensive coverage of peer-reviewed literature in engineering, science, and technology. The review focuses on research concerning deep learning-based human activity recognition (HAR) using FMCW radar. Search terms were grouped into three categories: application, sensor, and model. Each category includes keywords that broadly cover existing studies, as follows.

- Application: Human activity/action recognition*/classification*, fall detection*
- Sensor: FMCW/frequency modulated continuous wave radar*, *radar
- Model: deep learning, deep neural network (DNN), DNN*, CNN*, RNN*, long short-term memory (LSTM), LSTM*, autoencoder, convolutional autoencoder (CAE), few-shot learning, generative adversarial networks (GANs), transformer.

The asterisk (*) signifies random characters that facilitate flexible connecting with many keyword differences. Incorporating Boolean operators such as AND and OR into the search strings can enhance the search process. Search strategies are adapted to align with the characteristics of each database. In IEEE Xplore, the number of keywords in a search term is limited, so the following search string was used: ("Human activity recognition*" OR "fall detection*") AND ("*radar") AND ("deep learning"). For ScienceDirect, which does not support wildcards (*), the search string was adjusted to: ("Human activity recognition" OR "Human activity classification" OR "fall detection") AND ("radar" OR "FMCW radar") AND ("deep learning" OR "CNN" OR "DNN" OR "RNN" OR "LSTM" OR "AE" OR "few-shot learning" OR "GAN" OR "transformer" OR "convolutional autoencoder"). Lastly, in Springer Nature, where wildcards (*) are supported, the search terms were constructed to allow broader coverage: ("Human activity recognition*" OR "Human activity classification*" OR "fall detection*") AND ("*radar" OR "FMCW radar*") AND ("deep learning*" OR "CNN*" OR "RNN*" OR "LSTM*" OR "GAN*" OR "transformer*"). The final search for the results in this article was conducted on 30 July 2025.

As a result, we have identified a total of 2,164 records. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33] framework was employed to ensure transparency and objectivity. Fig. 1 details the selection process. The first step retained only journal articles, excluding reviews, conference papers, books, and non-English sources, resulting in 744 records. Of these, 644 articles published between 2018 and 2025 were retained, reflecting the period in which deep learning architectures began to dominate HAR research. Subsequent title, abstract, and full-text screening yielded 82 final articles. Exclusion criteria were applied to remove studies outside the scope of FMCW radar and deep learning, including the below.

- Application: monitoring of heart rate, outdoor, gesture recognition/classification, monitoring of vital signs, motion tracking of the head, estimation of pose, recognition of facial expression, monitoring of contactless vital signs, human gait, detection of targets, vehicle and tracking, and object tracking.
- Sensors: ultra-wideband (UWB) radar, pulse radar, continuous-wave (CW) radar, Doppler radar, stepped frequency continuous wave (SFCW) radar, wearable,

TABLE I: Comparative Summary of State-of-the-Art Surveys on Radar-Based Human Activity Recognition (HAR).

Study	Year	Scope / Primary Focus	Radar Modalities Covered	Identified Limitations / Gaps
Islam et al. [25]	2022	CNN-based HAR across multimodal data (smartphone, radar, vision)	Doppler, FMCW, UWB	Broad multimodal scope; lacks in-depth analysis of FMCW-specific DL pipelines.
Ullmann et al. [29]	2023	Continuous HAR with radar	FMCW, Doppler, Pulsed, mmWave	Focus on continuous recognition; limited discussion of DL architectures.
Zhang et al. [28]	2023	mmWave sensing technologies (tracking, motion, biometrics, imaging)	FMCW, Pulsed, mmWave	Application-oriented; no systematic review of DL models for HAR.
Hu et al. [26]	2023	Radar-based fall detection in healthcare	FMCW, UWB, SFCW	Narrow focus on fall detection; limited general HAR coverage.
Ahmed et al. [30]	2023	ML in radar-based healthcare monitoring	FMCW, SFCW, Pulsed	Highlights ML/DL for vital signs; lacks comprehensive DL-HAR synthesis.
Miazek et al. [32]	2024	Human behavior analysis using radar, LiDAR, and video	mmWave	Emphasizes multimodal fusion; FMCW-specific HAR only briefly mentioned.
Nocera et al. [31]	2024	ML for radar-based physiological signal monitoring	UWB, Doppler, FMCW	Focus on physiological sensing; HAR analysis secondary.
Tewari et al. [27]	2024	State-of-the-art radar sensing for fall detection	Pulsed, Doppler, FMCW, UWB, SFCW	Concentrates on fall monitoring; does not address broader DL-based HAR challenges.

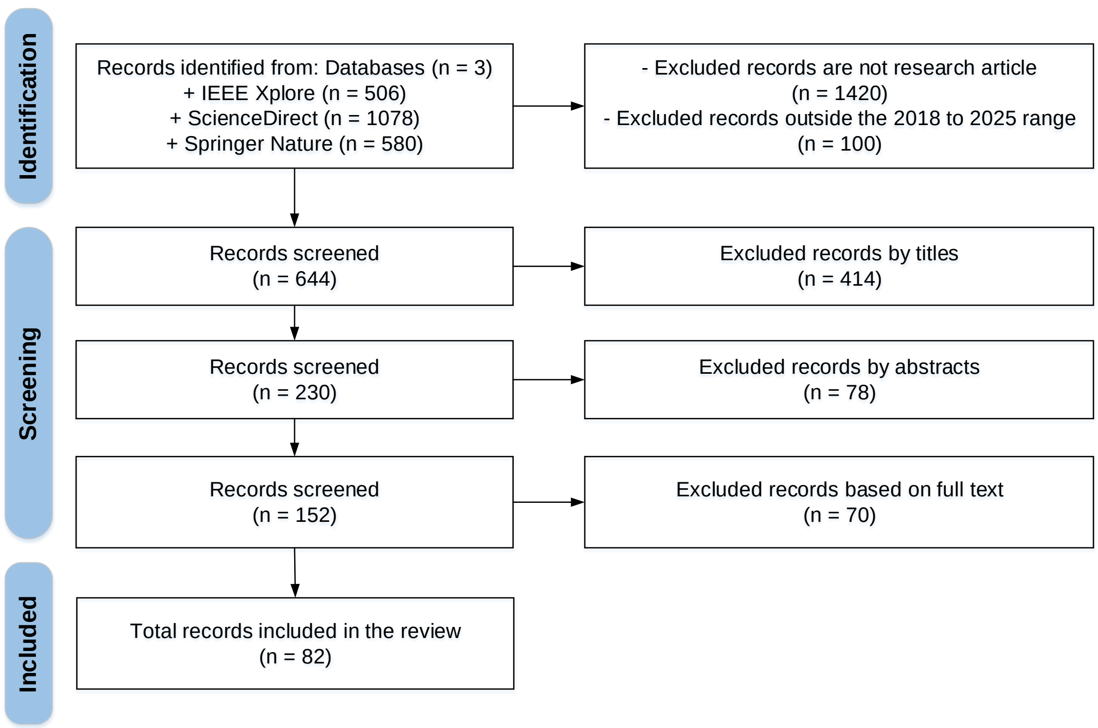


Fig. 1: Flow chart of the selection procedure of records according to PRISMA guidelines.

inertial measurement units, smartphone, camera, image and video, accelerometer, magnetometer, gyroscope, pyroelectric infrared (PIR), resolution thermal, wi-fi, and radio frequency (RF) sensing.

- Model: support vector machine, nearest neighbor, hidden Markov model, k-means clustering, decision tree, random forest, principal component analysis, and kNN.

These studies were considered at the scoping stage but excluded from in-depth analysis to maintain a focused review on FMCW radar with deep learning.

Finally, both bibliometric [34] and qualitative content analyses [35] were applied. Bibliometric analysis examined publication trends, citations, and keyword co-occurrence using VOSviewer [36] and Excel, while content analysis systematically identified themes including preprocessing strategies, dataset availability, deep learning architectures, and emerging challenges. This dual approach provides a balanced quantitative and qualitative synthesis of the field.

III. Results

A. What Is The State Of The Art On Deep Learning-Based HAR Using FMCW Radar?

The distribution of documents related to deep learning-based HAR using FMCW radar in the three databases surveyed from 2018 to 2025, as shown in Fig. 2. In 2018, no articles on this topic were found. From 2019 onward, publication activity grew steadily, rising from a single article in 2019 to 20 articles in 2023, 19 in 2024, and 14 in the first seven months of 2025. Most contributions appeared in IEEE Xplore-indexed journals, underscoring the field's strong alignment with the radar and signal-processing community. This trajectory illustrates not only quantitative growth but also a consolidation of research interest, marking HAR with FMCW radar as an emerging and increasingly competitive subdomain. The bibliometric analysis indicates that the articles are distributed among 22 journals, as illustrated in Fig. 3. The *IEEE Sensors Journal* dominates with 34 publications, followed by the *IEEE Internet of Things Journal* (11), the *IEEE Transactions on Aerospace and Electronic Systems* (6), and *Neural Computing and Applications* (3). Smaller contributions are scattered across a wide range of journals, with five journals having two articles and 11 journals having one. When mapped against journal impact metrics ("Scimago Journal & Country Rank" (SJR) and Journal Citation Reports(JCR)), the analysis shows that almost all papers were published in Q1 or Q2 journals (Fig. 4). The absence of the Q3, Q4 venues indicates that the technical rigor and perceived importance of this line of research are relatively high. Notably, the number of Q1 publications has steadily increased over time. That signals growing recognition of radar-based HAR in prestigious and high-ranking journals.

Citation analysis results are in Table III, which shows the top ten articles with the most citations. The article entitled "Continuous Human Activity Classification From FMCW Radar With Bi-LSTM Networks" [37] published in the IEEE Sensors Journal by A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, leading with 240 citations in Google Scholar. Other notable works, such as "mmFall: Fall Detection Using 4-D mmWave Radar and a Hybrid Variational RNN AutoEncoder" [38], "Noninvasive Human Activity Recognition Using Millimeter-Wave Radar" [39] and "Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures" [40] have also received substantial attention. This indicates that early methodological breakthroughs continue to anchor the field. In contrast, most 2024–2025 works remain lightly cited, a lag that reflects both their recency.

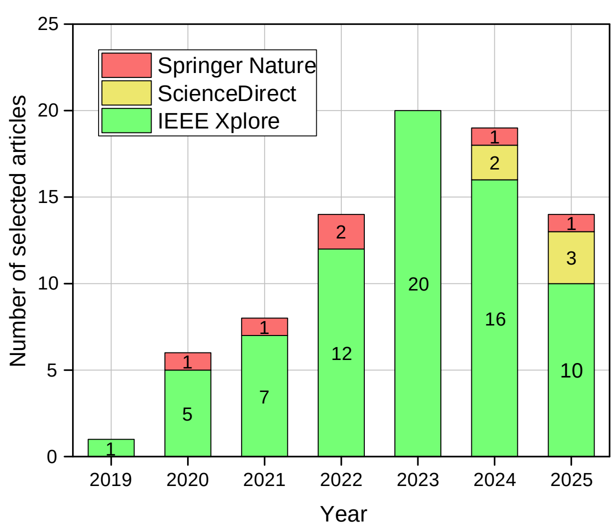


Fig. 2: Number of articles published during 2019-2025.

TABLE II: Top 10 most popular keywords.

Keyword	Occurrence
Human activity recognition	51
FMCW radar	39
Deep learning	25
Fall detection	21
Micro-Doppler	16
MmWave radar	15
Convolutional neural network	14
Feaature fusion	5
Long short-term memory	4
Point cloud	4

We also did keyword co-occurrence analysis using the VOSviewer software. We analyzed 82 articles to determine the popular keywords extracted from the articles. Fig. 5 illustrates 18 popular keywords after filtering out those with a frequency of less than 3 from a total of 190 keywords. The evolution of the research field can be observed through the progression of popular keywords over time. From 2018 to 2022, the studies were published throughout this period, focused on keywords such as "convolutional neural network", "semisupervised learning", "transfer learning", and "deep neural networks". Between 2022 and 2023, the keywords transferred to "long short-term memory", "machine learning", "range Doppler", "micro Doppler", "mmWave radar", "fall detection", "few-shot learning", and "human activity recognition". From 2023 to 2024, the keywords highlighted are "data augmentation", "point cloud", "FMCW radar", and "deep learning". From 2024 to 30 July 2025, studies began to emphasize keywords such as "feature fusion" and "lightweight network". Table II presents the top 10 most popular keywords among 190 keywords of the analyzed articles. These include "human activity recognition" (51 occurrences), "FMCW radar" (39 occurrences), "deep learning" (25 occurrences), "fall detection" (21 occurrences), and "micro-Doppler" (16 occurrences). Ranking 9th and 10th are "long short-term memory" and "point cloud", each with 4 occurrences.

Overall, the bibliometric evidence reveals three clear trajectories: (i) rapid quantitative growth, (ii) concentration of publications in high-quality engineering journals. Together, these findings confirm FMCW radar-based HAR as a young but emerging field with significant technical innovations.

B. What Techniques Are Used To Process HAR's FMCW Radar Signals?

1) Principles and architectures of FMCW radar sensor: A complete FMCW radar includes: (1) frequency synthesizer, (2) transmitting antenna (Tx), (3) receiving antenna (Rx), (4) frequency mixer, (5) low-pass filter (LPF), (6) analog-to-digital converters (ADCs), and (7) microprocessor, as shown in Fig. 6a. The transmitted frequency-modulated chirp is

$$S_{Tx}(t) = A_{Tx} \sin\left(2\pi \left(f_0 t + \frac{St^2}{2}\right)\right), \tag{1}$$

where f_0 is the start frequency and S is the chirp slope. The received signal $S_{Rx}(t)$ is a delayed version of $S_{Tx}(t)$

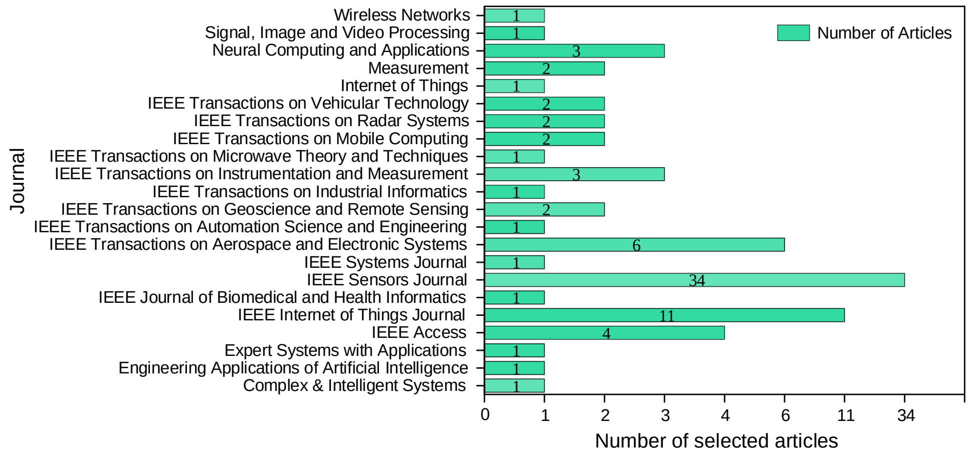


Fig. 3: Journals with the number of published articles.

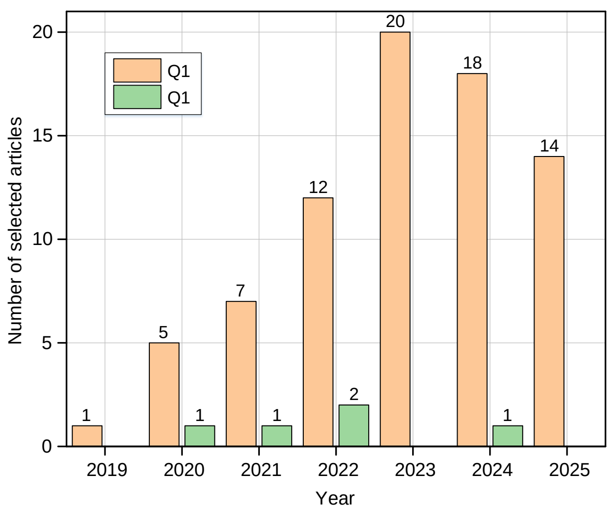


Fig. 4: Distribution of papers following journal ranking.

TABLE III: Top 10 articles with the most number of citations.

Article Title	Citation
Continuous Human Activity Classification From FMCW Radar With Bi-LSTM Networks [37]	240
mmFall: Fall Detection Using 4-D mmWave Radar and a Hybrid Variational RNN AutoEncoder [38]	124
Noninvasive Human Activity Recognition Using Millimeter-Wave Radar [39]	122
Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures [40]	120
Human Activity Classification Based on Point Clouds Measured by Millimeter Wave MIMO Radar With Deep Recurrent Neural Networks [41]	104
Deep Learning Radar Design for Breathing and Fall Detection [42]	97
Radar-Based Human Activity Recognition Using Hybrid Neural Network Model With Multidomain Fusion [43]	80
A Millimetre-Wave Radar-Based Fall Detection Method Using Line Kernel Convolutional Neural Network [44]	79
Activity Classification Based on Feature Fusion of FMCW Radar Human Motion Micro-Doppler Signatures [45]	79
Fall Detection System Using Millimeter-Wave Radar Based on Neural Network and Information Fusion [46]	76

due to the target range d, with delay $\tau = \frac{2d}{c}$, shown in Fig. 6b. Mixing the signals $S_{Tx}(t)$ and $S_{Rx}(t)$ and passing them through the LPF produces the intermediate frequency (IF) signal, which encodes information about the target range, velocity, and angle.

FMCW radar architectures are broadly classified into two types based on mixer and baseband implementations: real mixer with real baseband (RMRB, marked in a green rectangle) and quadrature mixer with complex baseband (QMCB, marked in a red rectangle), as shown in Fig. 6a. The RMRB architecture uses a real mixer whose IF output, after low-pass filtering, is expressed as:

$$S_{IF}(t) = A_{IF} \cos\left(2\pi S\tau t + 2\pi f_0 \tau - \pi S\tau^2\right), \quad (2)$$

where $A_{IF} = \frac{A_{Rx}A_{Tx}}{2}$ and the beat frequency $f_b = S\tau$ is proportional to the target range. In contrast, the QMCB architecture employs both a real and a quadrature mixer, producing in-phase (I) and quadrature (Q) components. The complex IF signal is given by

$$S_{IF}(t) = A_{IF} e^{-j(2\pi S\tau t + 2\pi f_0 \tau - \pi S\tau^2)}, \quad (3)$$

The ADC samples the IF signal to collect raw radar data and then processes these digital signals through a microprocessor. This complex baseband processing enables more accurate estimation of target parameters, improving the system's resolution.

2) Range, velocity, and angle estimation: The target range estimation is derived from the beat frequency f_b of the IF signal generated by a target at distance d, expressed as $f_b = \frac{S2d}{c}$. Consequently, the target range can be computed by:

$$d = \frac{f_b c T_c}{2B}. \quad (4)$$

Multiple targets produce multiple peaks in the IF signal spectrum identified via a fast Fourier transform (range-FFT). The minimum distinguishable distance between two targets, or range resolution d_{res}, is inversely proportional to bandwidth:

$$d_{res} = \frac{c}{2B} \quad (5)$$

Velocity estimation exploits the phase difference $\Delta\theta = \varphi_{IF2} - \varphi_{IF1}$ between consecutive chirps. A stationary target exhibits no phase change ($\Delta\theta = 0$), while a moving

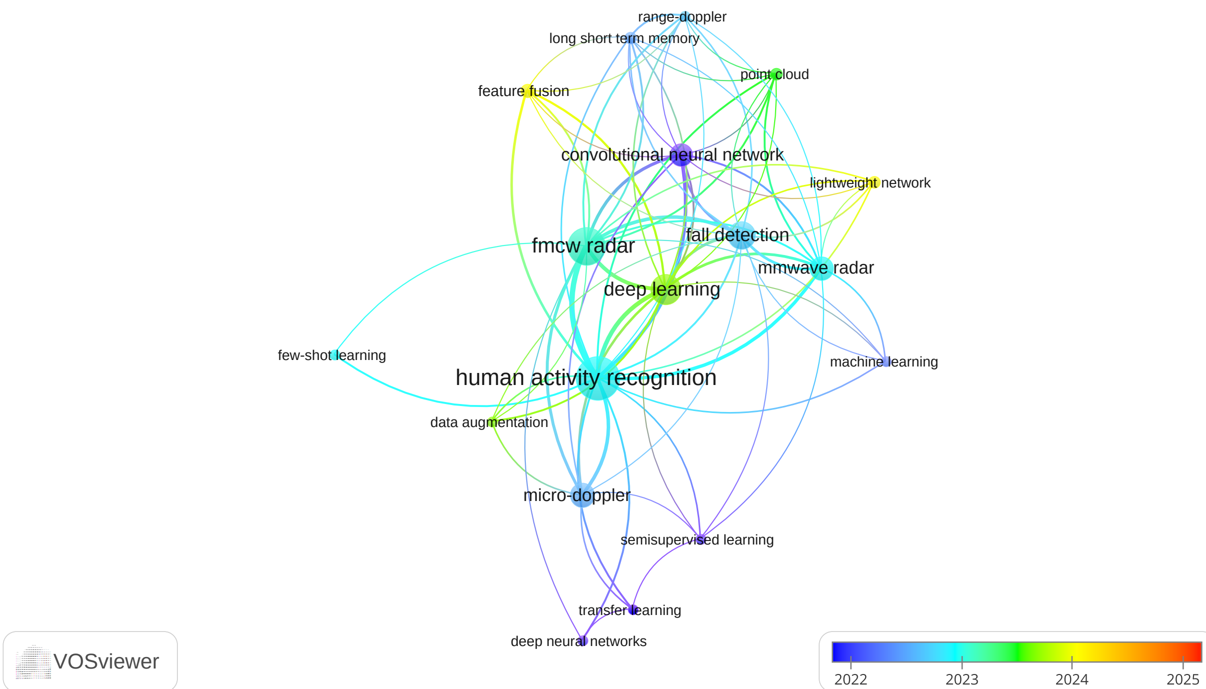
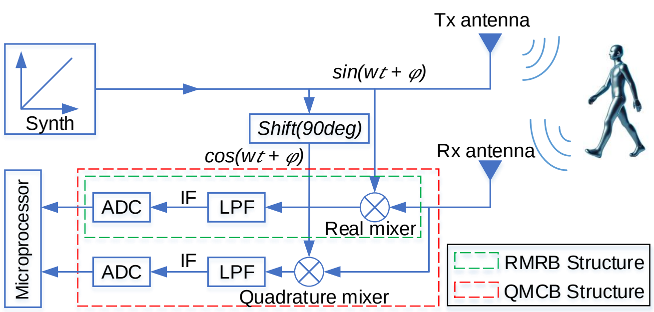
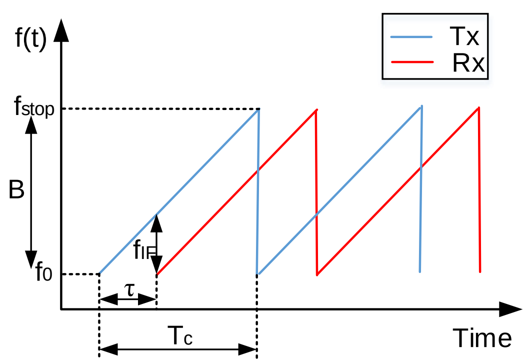


Fig. 5: The popular keywords related to deep learning-based HAR using FMCW radar.



(a)



(b)

Fig. 6: (a) FMCW radar sensor schematic block diagram. (b) Chirp signal.

target introduces a phase shift related to its velocity v by $\Delta\theta = 2\pi f_0 \Delta\tau$, where $\Delta\tau = \frac{2vT_c}{c}$. The velocity can then be calculated as:

$$v = \frac{\lambda \Delta\tau}{4\pi T_c} \tag{6}$$

with λ is the radar wavelength and $f_0 = \frac{c}{\lambda}$ the carrier frequency. Doppler-FFT applied over N chirps enables differentiation of targets with velocity resolution greater than $v_{res} = \frac{\lambda}{2T_f}$, where T_f is the total observation time.

Angle of arrival (AoA) estimation necessitates multiple receive antennas (at least two receive antennas), commonly realized via multiple-input multiple-output (MIMO) radar architectures, to resolve target azimuth and elevation. The phase progression across M receiving antennas is as follows:

$$[0, 2\pi\frac{lsin(\alpha)}{\lambda}, 4\pi\frac{lsin(\alpha)}{\lambda}, ..., 2\pi(M-1)\frac{lsin(\alpha)}{\lambda}] \tag{7}$$

where l is the antenna spacing and α is the AoA. The AoA is estimated from the phase difference between adjacent antennas $\Delta\phi = \frac{2\pi}{\lambda} lsin(\alpha)$. This change is determined through a third FFT, known as angle-FFT.

3) Techniques for processing FMCW radar signals: The overall processing pipeline incorporates range-FFT to extract beat frequencies for range-time, Doppler-FFT across chirps for range-Doppler, and angle-FFT for angular information, as depicted in Fig. 7. Furthermore, the short-time Fourier transform (STFT) is utilized across the chirps to analyze the phase shift and generate the micro-Doppler signature. Leveraging FFTs in both the horizontal and vertical orientations of virtual receiving antennas facilitates the precise determination of the target's azimuth and elevation angles. From the resulting range-angle representation, various filtering and clustering algorithms can

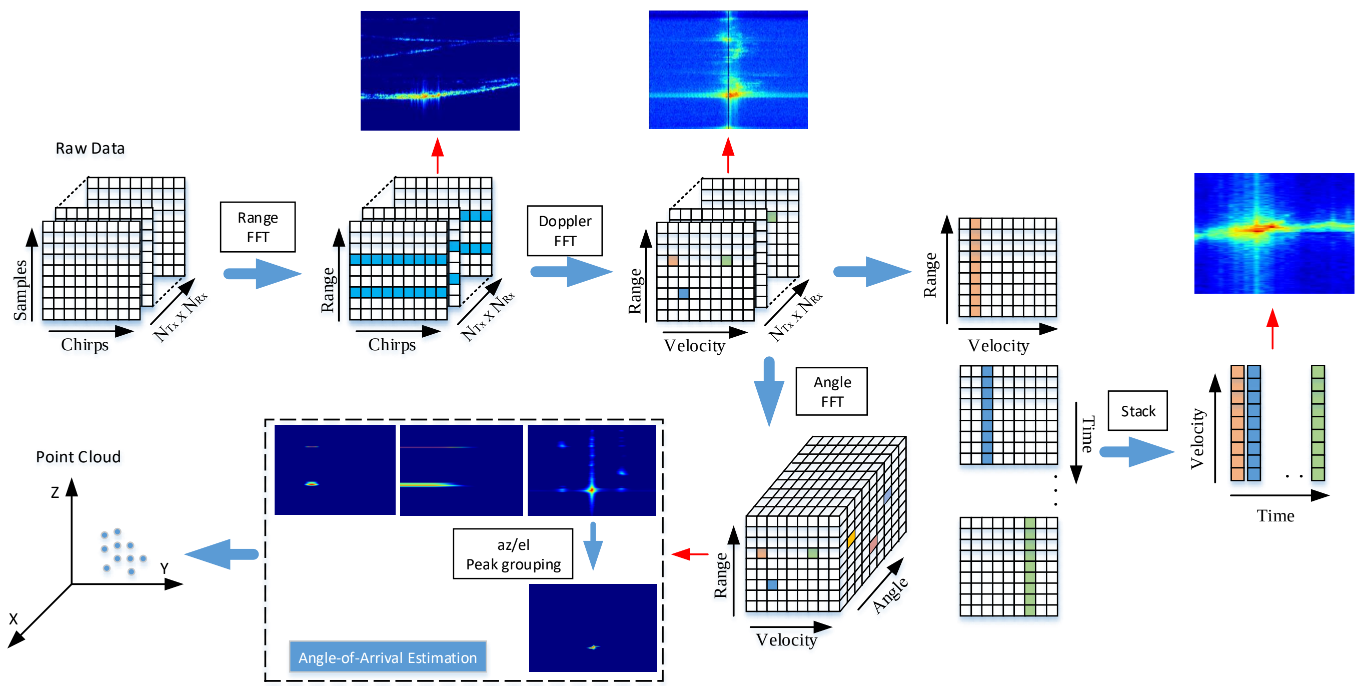


Fig. 7: General signal processing pipeline for radar signals and common feature representation domains used for HAR

be applied to generate a sparse point cloud that reflects the target's information. Table IV shows a comprehensive list of the features that were used in DL-based HAR using FMCW radar sensors.

Among the analyzed publications, micro-Doppler is identified as the most common feature, appearing in 41 articles [37], [40], [42], [45], [47]–[83]. Its popularity stems from its robust capability to capture body and limb motion characteristics via velocity information, enabling the differentiation of various human activities. Additionally, features like point clouds [38], [39], [41], [84]–[91], and range-Doppler [92]–[98] are extensively utilized, providing supplementary spatial, distance, and velocity information. Conversely, deep learning models rarely use raw data [44] due to the difficulties in feature extraction. Recent advancements in the field indicate an increasing trend toward integrating multi-dimensional data to enhance recognition performance. Examples include combinations like (micro-Doppler, point clouds) [69] and (range-Doppler, point clouds) [99], [100], which merge spatial and temporal data for a deep feature. More advanced formats, such as (Range-Doppler-Time) [101]–[104], (Range-Time, Range-Doppler, Doppler-Time) [43], [105], [106], (micro-Doppler, range-time, azimuth-time, elevation-time) [107] and (range-time, Doppler-time, angle-time) [108], have gained increasing attention, reflecting efforts to leverage complex, multi-dimensional features to enhance accuracy and robustness. These innovations reflect the evolution of deep learning models, with a growing emphasis on utilizing abundant data representations to advance recognition capabilities and generalize HAR systems for diverse environments.

TABLE IV: The features of FMCW radar signal.

Features	References
micro-Doppler	[37], [40], [42], [45], [47]–[83]
micro-Doppler, Point clouds	[69]
micro-Doppler, Range-Doppler-Time	[109], [110]
micro-Doppler, Range-Time, Azimuth-Time, Elevation-Time	[107]
Point clouds	[38], [39], [41], [84]–[91]
Range-Doppler	[92]–[98]
Range-Doppler, micro-Doppler	[111]
Range-Doppler, Point clouds	[99], [100]
Range-Doppler, Range-Angle	[112]
Range-Doppler, Range-Doppler-Time	[113]
Range-Doppler-Time	[101]–[104]
Range–Horizontal Angle, Range–Vertical Angle, Range-Doppler	[46]
Range-Time	[114]
Range-Time, Doppler-Time	[115]–[117]
Range-Time, Doppler-Time, Angle-Time	[108]
Range-Time, Range-Doppler, Doppler-Time	[43], [105], [106]
Raw data	[44]

C. How Have Deep Learning Models Evolved For FMCW Radar HAR?

The evolution of deep learning models for FMCW radar HAR has progressed through successive architectures and data representations, with each stage targeting specific challenges in accuracy, robustness, and deployment feasibility, as summarized in Fig. 8. In the initial phase (2019–2022), CNN-based approaches dominated, focusing on extracting spatial–spectral features from micro-Doppler and range–Doppler spectrograms. From 2020 to 2023, hybrid and multi-domain fusion models integrating range, Doppler, time, angle, and even radar point-cloud data became increasingly important.

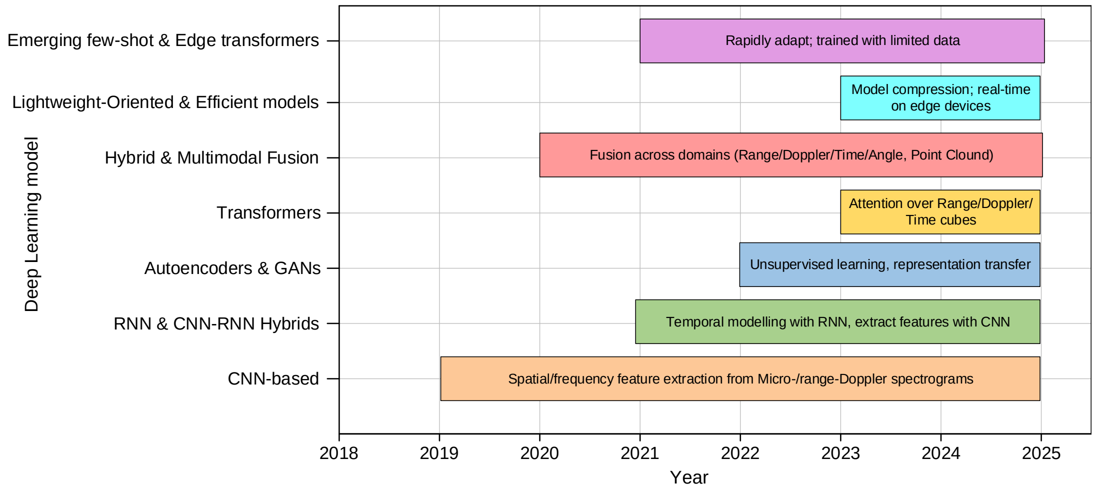


Fig. 8: Evolution of deep learning models for FMCW radar HAR (2018-2025).

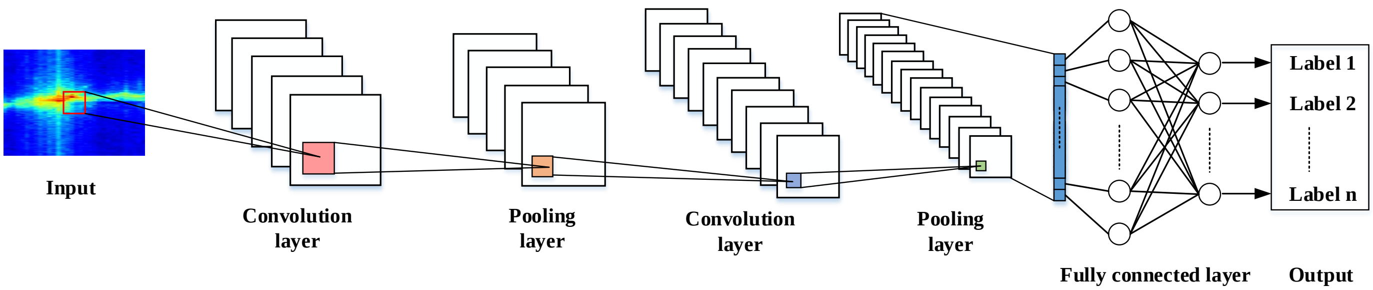


Fig. 9: General architecture of CNN-based models.

Between 2021 and 2023, researchers integrated RNNs and CNN–RNN hybrids to capture temporal dynamics while leveraging CNNs for local feature extraction. Additionally, few-shot frameworks and transformer variations capable of fast adaptation and training with minimal data have been developed. From 2022 to 2024, autoencoder-based methods gained prominence, particularly for unsupervised representation learning and cross-domain feature transfer. Finally, between 2023 and 2025, transformer architectures emerged, applying attention mechanisms across range–Doppler–time cubes to model long-range dependencies. In parallel, lightweight-oriented models for low-latency, energy-efficient devices have emerged. They are anticipated to shape future research toward real-time, resource-constrained HAR applications.

1) CNN-based models: Convolutional Neural Networks (CNNs) have become the dominant deep learning architecture in FMCW radar-based HAR, particularly due to their efficiency in extracting discriminative spatial–spectral features from radar representations. CNNs leverage weight sharing across convolutional kernels, which not only reduces parameter count but also enables efficient learning of local feature hierarchies. This property is especially advantageous for IoT and edge deployments, where memory and computation budgets are limited.

In the radar-based HAR task, raw radar signals are commonly transformed into 2D time–frequency or range–Doppler maps to facilitate spatial–spectral feature learning, making CNNs a natural fit for this domain [45], [56], [61], [63], [94], [115]. As shown in Fig. 9, the CNN pipeline typically consists of convolutional and pooling layers for hierarchical feature extraction, followed by fully connected layers for activity classification. Most related research utilizes micro-Doppler (MD) representations as CNN inputs, where MD maps are treated as 2D images to be analyzed for human motion patterns. Although CNNs achieve high accuracy, their direct deployment on IoT devices remains challenging due to resource and latency constraints. Several works have therefore focused on lightweight CNN variants, using grouped convolutions [64] or strip pooling [62] to reduce complexity without sacrificing performance.

Additionally, many articles have focused on extending CNNs for multi-branch and multi-view feature learning. In [111], a 3D-CNN applied to range–Doppler data produced superior results compared to traditional 2D-CNNs

Kim et al. [102] considered the time-Doppler map for each range bin and introduced a range-distributed CNN with multiple branches designed to analyze each range-time-Doppler map explicitly. Alternatively, Li et al. [61] presented a multi-attention fusion CNN combining three branches with different feature attention mechanisms on the micro-Doppler map; and the results are combined through ensemble learning for omnidirectional HAR. Similarly, mmCMD [60] visualized continuous motion signatures as images and applied YOLOv5 for activity detection.

Beyond micro-Doppler and range-Doppler maps, radar point clouds have also emerged as a promising input modality.

Yu et al. [39] proposed a dual-view CNN that processes voxelized point clouds from two perspectives (XoZ and YoZ), demonstrating robust cross-domain generalization. Rezaei et al. [84] further confirmed the feasibility of CNNs with point cloud inputs for unobtrusive HAR. Table V summarizes the performance of CNN-based methods along with the corresponding experimental data details.

The table highlights several important trends in the use of CNNs for FMCW radar-based HAR. First, CNN-based methods consistently achieve high classification accuracy, often exceeding 90–99% across different datasets, confirming their strong ability to capture discriminative spatial features. However, such performance is frequently tied to specific datasets or controlled environments, raising concerns about generalization in more complex real-world scenarios and emphasizing the need for cross-dataset validation. A second observation is the evolution of input representations: while early work predominantly relied on micro-Doppler and range-Doppler spectrograms, more recent studies have moved toward richer modalities such as radar point clouds and multi-view representations. This trend indicates that the choice of feature representation is as crucial as the depth of the network, particularly for capturing fine-grained human motions.

Another clear development is the shift from conventional single-stream CNNs (e.g., AlexNet, VGG) to multi-branch and attention-enhanced CNNs, which are designed to process heterogeneous radar features in parallel and integrate them through specialized fusion strategies. Such architectures consistently outperform simple baselines, suggesting that domain-specific adaptations are essential for robust radar HAR. At the same time, recent efforts have begun to address the computational limitations of CNNs in edge and IoT scenarios. Studies leveraging grouped convolutions, lightweight networks such as MobileNet, or hardware accelerators demonstrate the feasibility of balancing accuracy with efficiency, though such works remain limited compared to accuracy-focused designs. Finally, despite the consistently strong recognition performance reported, the lack of standardized benchmarks across radar HAR studies complicates fair comparison. Differences in radar configurations, activity definitions, and dataset sizes make it difficult to isolate the contributions of model design from experimental conditions, pointing to an urgent need for unified evaluation protocols.

From these observations, they suggest that CNNs remain a cornerstone of radar-based HAR, but their future development will likely hinge on improved generalization across domains, efficient adaptation to edge devices, and systematic benchmarking. To address temporal dynamics more explicitly, many studies have further extended CNN-based representations with sequence modeling architectures such as recurrent neural networks (RNNs), which will be discussed in the following subsection.

2) RNN and CNN-RNN Hybrids models: Recurrent Neural Networks (RNNs) are widely used for sequential data processing because they retain information from previous time steps through a hidden state mechanism [120]. This temporal memory enables HAR systems to capture motion dynamics across radar frames, complementing the spatial features often

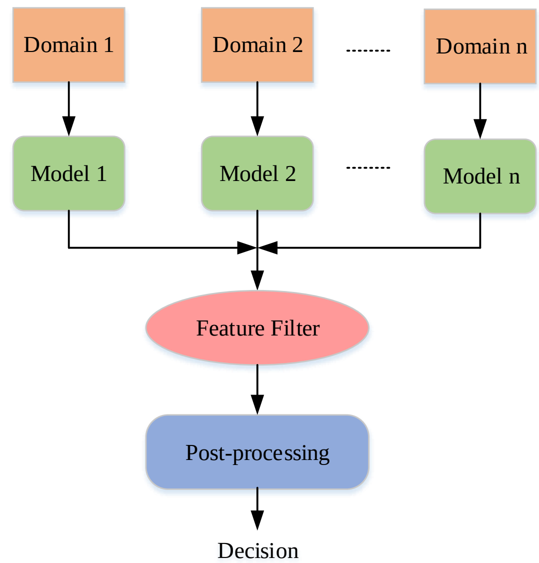


Fig. 10: Fundamental architecture of multimodal fusion models.

emphasized by CNNs. However, conventional RNNs suffer from vanishing and exploding gradient issues, which limit their ability to model long-range dependencies. To address these limitations, Long Short-Term Memory (LSTM) networks were introduced [121], incorporating gating mechanisms to preserve long-term context while filtering irrelevant short-term information. For example, Kittiyanpunya et al. [100] applied an LSTM network for fall detection using 1-D Doppler velocity and z-axis point cloud features. By introducing a sliding-window mechanism to mitigate spurious detections, the system achieved an accuracy of 99.5%. This demonstrates the potential of LSTMs for radar-based HAR, although their reliance on carefully engineered sequential inputs limits scalability across heterogeneous radar settings.

Simplified variants such as Gated Recurrent Units (GRUs) [122] further improve computational efficiency, making them suitable for real-time deployment. Abedi et al. [97] developed an autonomous in-home gait monitoring and activity recognition system using GRUs with joint time-frequency inputs, enabling deployment on IoT hardware. While these works illustrate the feasibility of standalone RNNs, the broader radar HAR literature rarely adopts them in isolation, as spatial feature extraction from radar images is typically indispensable. Table VI summarizes representative RNN-only studies.

Given these limitations, hybrid architectures that combine CNNs with RNNs have become prevalent. The rationale is that CNNs are adept at capturing local spectral or spatial features from radar representations (e.g., spectrograms, range–Doppler maps), meanwhile, RNNs excel at modeling temporal correlations across frames. In [41], a CNN first extracted pose features from point cloud frames, which were then processed by an LSTM to model inter-frame dynamics. This hybrid model outperformed the CNN-only baseline, improving recognition accuracy from 96.1% to 98.3%. Similarly, Li et al. [52] integrated ResNet18 with a Bi-LSTM enhanced by attention

TABLE V: Summary of studies utilizing CNN-based models.

Study	Experimental Data	Performance Details
[78] (2019)	Each person performs the action 3 meters in front of the radar, 9 single persons, 4 designed activities in place, a total of 11,520 samples.	Investigates CNN acceleration on FPGA for human activity classification. Achieves 30.42% speedup compared to GPU implementation.
[67] (2021)	Glasgow dataset [118]	97.58% classification accuracy, utilizing a three-branch CNN model with 1-D color channels of micro-Doppler images.
[56] (2022)	Radar mounted on the ceiling, 2.5 m above the floor. The experimental area is placed with some experimental tables, chairs, a bed, and sundries12 subjects, 29 designed activities, a total of 192,400 samples.	95.88% accuracy for fall detection using VGGNet with filtered DTmap through pattern contour-confined.
[115] (2022)	Experiment setup similar to [56]	95.48% accuracy for sudden and soft fall detection using the CNN model.
[45] (2022)	Glasgow dataset	Classification accuracy of 99.77% using VGG+Alexnet.
[111] (2022)	PARrad dataset [111]	Achieved 92% and 95% accuracy in activity classification for the CNN-MD network and the CNN-RD network.
[102] (2022)	Glasgow dataset	The mean classification accuracy is 95.71% through range information using multi-branch CNN.
[63] (2022)	Two radars are used to simultaneously collect data at a height of 1 m, 6 subjects, 7 designed activities, a total of 1,680 samples.	The recognition accuracy of 94.14% using a CNN model from Stockwell transformed signals of two radars.
[39] (2022)	The radar was placed at a height of 2.9 m in a 2 × 2.5 m room, 4 subjects, 7 designed activities in multiple orientations, a total of 120,000 samples	Using point cloud into CNN model, achieve 97.61% and 98% accuracies of fall detection and activity classification.
[60] (2023)	Radar mounted on the ceiling, 1.8 m above the floor with a pitch angle range from $-20°$ to $20°$ and an azimuth angle range from $-50°$ to $50°$ in a 4×4.8 m test area was selected within an 8×9 m indoor hall, participating volunteers performed 12 actions, total 40,560 samples.	The mmCMD approach utilizes micro-Doppler signatures as images into YOLOv5, achieving a mean average precision of 93%.
[84] (2023)	Radar mounted on the ceiling 3 m above the floor with a field of view 120° in both azimuth and elevation directions. Experiments were conducted in an area of 10.4 × 10.4 m^2, 2 subjects, 9 designed activities, a total of 154,040 samples in trial 1 and 143,333 samples in trial 2	CNN-based deep learning model using point cloud achieved an accuracy of 92.3%.
[94] (2024)	Radar positioned near the ceiling at a height of 3 m, the participation of 10 males performing 5 activities within a 5 m range from the radar stand, totaling 4,620 samples.	RDTNet model achieves an accuracy of 99.71% in fall detection using the range-Doppler map and range-Doppler cube input data.
[61] (2024)	Radar was set at a height of 1.5 m with a 15° of inclination, 7 designed activities within a range of 1 m to 12 m from the radar, totaling 2,100 samples.	MAEF-Net uses a three-branch CNN with filtered micro-Doppler, achieving 98.3% accuracy for addressing omnidirectional HAR.
[62] (2024)	Experiments were conducted in a rectangular area of 5 x 4 m, with the radar placed at a height of 1.2 m above the floor, 10 subjects, 6 designed activities, a total of 4,120 samples.	99.28% classification accuracy using the CNN model with the time-Doppler map input data.
[64] (2024)	Radar is placed against a wall at a height of 1.9 m and 2 m above the floor in scenes one and two, 7 subjects, 12 designed activities, total 1,716 samples.	The recognition accuracy is 95.79%, utilizing a lightweight CNN model with micro-Doppler signature input.
[51] (2024)	Glasgow dataset	95.83% classification accuracy using ResNet 18 to classify 6 different activities.
[81] (2025)	The radar is mounted on a fixed frame at a height of 1.2 m above the ground, 12 volunteers aged 19 to 38 performed 5 activities, a total of 60,000 samples.	The AB-TCN method's classification accuracy improves by over 2% compared to the CNN-LSTM.
[76] (2025)	Human Activity Data with a 5.8 GHz FMCW Radar [99].	MobileNetV2, GoogLeNet, AlexNet, VGG16, and VGG19 are used to classify various activity classes based on the fourth-order Fourier synchrosqueezing transform (FSST4) technique and time–frequency representation. AlexNet achieved the highest accuracy of 99.40%.
[83] (2025)	Glasgow dataset, CI4R [119], I+ Lab RadSet v1 [62].	The MC-GCLU network consists of the masked concatenation module and the gated convolution and linear unit module, and it utilizes MD signatures as input for classifying activities with 97.4% accuracy on the RadarMD dataset.

TABLE VI: Summary of studies utilizing RNN-based models.

Study	Experimental Data	Performance Details
[90] (2022)	The experiments were carried out in a room with a 3 × 3 m space with four subjects and four classes designed for activities, a total of 1011 data files.	Used LSTM and DBSCAN for clustering point cloud data to classify activities. It achieved a recognition accuracy over 95%.
[100] (2023)	The experiments were divided into five categories in different rooms, and radar was mounted on a 1.6 m height tripod; 10 individuals participated and performed 9 different activities, a total of 270 sample sequences.	LSTM network is used to classify 9 activities fed by a 1-D point cloud in the z-axis and Doppler. The main focus is the detection of falls with an accuracy of 96.90%.
[97] (2023)	A total of 310,357 samples of the range-Doppler map were generated from seven participants and six different activities	The overall precision of the GRU model for classifying activities is 93%.

TABLE VII: Summary of studies utilizing CNN-RNN hybrid models.

Study	Experimental Data	Performance Details
[41] (2021)	Used 19 subjects, 7 designed activities, a total of 160,930 samples.	CNN+LSTM are applied to classify seven different activities, achieving over 97% classification accuracy by analyzing point cloud variations.
[52] (2023)	The radar is placed on a tripod with a vertical distance of 1.2 meters from the ground and at a horizontal distance of 2 meters from the subject, 13 subjects, 5 designed activities, and 7,680 samples.	Resnet18+Bi-LSTM is used to classify five different activities with micro-Doppler input data. It achieved a recognition accuracy rate of 95.79%.
[95] (2023)	The dataset was collected in two different environments, 5 subjects, 5 designed activities, totaling 500 samples.	CNN + LSTM on range Doppler maps to classify different activities, achieving 93.2% accuracy for five actions and 96.8% for fall vs. non-fall classification, with a 1.04 energy precision ratio on an edge device.
[103] (2024)	The radar height is set to 0.9 m, 12 subjects, 7 designed activities, totaling 1,948 samples.	Anti-fixed interference algorithm for range Doppler maps cancellation used to train CNN + Bi-LSTM for fall detection, achieving precision rates of 96.16%, 94.21%, and 93.06% in various interference scenarios.
[82] (2024)	The indoor experiments are set in a 7 × 5 m indoor laboratory with radar set at 1.1 m in height, 12 participants, and 6 designed activities, totaling 6,000 samples.	The CCLN model used CNN-LSTM to extract spatial and temporal features from micro-Doppler spectrograms. CCLN achieves an average accuracy of 95.91% in activity recognition, and using the AAT filter improves the accuracy by 1.14%.
[86] (2025)	The dataset was collected in two indoor environments and with radar at heights of 1.2 m and 2 m above the ground for four experimental setups. In the experiment, 15 subjects performed 9 actions in turn, totaling 64800 samples.	Utilizing triple-view CNNs together with LSTM models for point cloud input achieved an average accuracy for classification of almost 99.63% across nine human activities.

mechanisms, enabling selective focus on salient temporal cues. This design achieved 99.86% accuracy on a self-collected dataset, surpassing conventional CNN-LSTM pipelines.

Subsequent studies have refined the hybrid paradigm to enhance robustness and efficiency. Gianoglio et al. [95] validated CNN-LSTM performance on range–Doppler maps in edge scenarios, balancing accuracy (up to 96.8% for fall detection) with low energy consumption. Ma et al. [103] incorporated interference-cancellation preprocessing, maintaining over 94% precision across challenging environments. Recent work has further extended hybrid models with triple-view CNNs [86] and adaptive filtering [82], consistently reporting accuracies exceeding 95%. Table VII summarizes representative CNN-RNN hybrid models.

In summary, standalone RNNs remain useful in lightweight, real-time HAR systems, particularly when hardware constraints necessitate low model complexity. However, CNN-RNN hybrids have become widespread since they effectively integrate spatial and temporal representations. The integration is capable of addressing the multi-dimensional complexity inherent in radar-based HAR.

3) Autoencoder and GANs-based models: Autoencoders (AEs) and generative adversarial networks (GANs), which are generative models, have been explored in FMCW radar HAR to address two complementary challenges. AEs primarily target feature representation, denoising, and interference mitigation, while GANs focus on data scarcity and augmentation.

AEs are neural networks that learn to reconstruct input data by compressing it into a lower-dimensional representation (encoding) and then reconstructing the original data from that compressed representation (decoding). In radar HAR, this capability has been leveraged both for representation transfer and for signal enhancement. In [59], an AE-based approach outperformed CNN-based transfer learning techniques on the University of Glasgow radar dataset by pre-training the AE to capture latent representations, after which the decoder was removed and the encoder fine-tuned in a supervised manner. By contrast, Raeis et al. [96] applied convolutional AEs for interference cancellation as a feature enhancement step, improving multi-person HAR but at the cost of retraining for each unseen scenario. This highlights the key limitation of AE-based approaches: their latent representations are highly task-dependent, which reduces generalizability. From a performance perspective, AE-based methods generally achieve 85–97% accuracy in single-person scenarios but drop sharply (to around 73%) in multi-person conditions (Table VIII), underscoring their sensitivity to interference and domain variability.

TABLE VIII: Summary of studies utilizing Autoencoder and GANs-based models.

Study	Experimental Data	Performance Details
[59] (2022)	Glasgow dataset	Autoencoder model achieves an accuracy of 88% for activities of daily living using the micro-Doppler input data.
[53] (2022)	Measurement dataset: The radar is placed on a table with a height of 1.2 m in a 6 x 6 m indoor laboratory, 30 subjects, 5 designed activities, totaling 900 samples; Public dataset: Glasgow dataset	Focuses on using a Wasserstein-refined generative adversarial network with gradient penalty (WRGAN-GP) to synthesize micro-Doppler spectrograms, enhancing training datasets for deep convolutional neural networks. For the measurement dataset, the average classification accuracy is 98.7% in the validation set, 94.9% in the test set.
[54] (2022)	The radar is placed on a table with a height of 1.5 m at a walkway that was 6 m long and 3 m wide, 10 subjects, 14 designed activities, totaling 1,400 samples	Propose a multibranch GAN (MBGAN) that integrates domain knowledge and physics-aware metrics to improve the synthesis of RF micro-Doppler signatures, achieving an accuracy of 89.23%.
[68] (2023)	Glasgow dataset	Semisupervised Triple-GANs are used to classify six typical human motions by leveraging micro-Doppler signatures and data augmentation techniques, achieving 90% accuracy with only 10% labeled data.
[96] (2024)	Experiments were carried out in the laboratory with a 4 × 10 m space with two participants and seven designed activities, a total of 3,455 samples.	CAE-MAS used a convolutional autoencoder, enabling accurate detection of multiple human activities. It achieves recognition accuracies of 97.13% for one human and 73.37% for two humans.
[85] (2025)	The experiments were conducted in two different experimental environments, and the radar was installed at a height of 1.8 m, covering a 4 m × 10 m detection area with five participants and eight designed activities in place, a total of 13,250 samples.	ABC-HF achieves 94% higher performance than other methods, such as OC-SVM, AE-MLP, and mmFall, in multiple human fall scenarios in indoor office spaces and industrial environments.
[80] (2025)	Glasgow dataset.	CRSFD used the variable information bottleneck technique, achieving the highest HAR accuracy, with more than 97% recognition accuracy using the time Doppler map input.

A natural solution to the lack of data problem is data augmentation, in which GANs are effective. GANs consist of two neural networks (i.e., the generator and the discriminator) that engage in adversarial training: the generator produces synthetic radar spectrograms, while the discriminator evaluates their realism. After convergence, the generator can produce large volumes of high-quality synthetic data for training. This makes GANs particularly attractive in radar HAR, where annotated datasets are scarce and expensive to collect. To address the instability of conventional GAN training, Qu et al. [53] introduced a refinement with gradient penalty (WRGAN-GP), achieving accuracies of 94.9–98.7% on both real and synthetic datasets. Rahman et al. [54] proposed a physics-aware multibranch GAN that integrated domain priors to generate more realistic micro-Doppler signatures, improving recognition robustness (89.2%). Liu et al. [68] developed a semi-supervised Triple-GAN framework that incorporated a classifier into the adversarial loop, allowing effective training with only 10% labeled data and achieving 90% accuracy. These studies illustrate how GANs not only augment data but also enable label-efficient learning strategies, which are particularly relevant for FMCW radar HAR where labeling effort is prohibitive. Across the studies, GAN-based approaches typically report higher robustness than AEs, with accuracies often exceeding 90–95% even under limited training data (Table VIII). However, they face challenges such as mode collapse, instability, and the risk that synthetic data may not generalize across domains.

Looking forward, hybrid models that combine the denoising and latent representation strengths of AEs with the data generation power of GANs (e.g., VAE-GANs), as well as diffusion-based generative models, are expected to play a central role. These could bridge the gap between reliable feature learning and scalable data augmentation.

4) Transformers models: Unlike traditional sequential models such as RNNs and LSTMs, transformers rely on the self-attention mechanism, which enables parallel input processing and context-aware feature extraction across long sequences. This design not only improves computational efficiency but also allows transformers to capture complex dependencies within FMCW radar data. Their ability to selectively attend to relevant information across time, frequency, and spatial domains makes them highly suited for human activity recognition, particularly when range, Doppler, and angular features must be modeled jointly.

Recent work has primarily adopted vision-transformer (ViT) variants, often in combination with CNNs for low-level feature extraction. Gu et al. [57] integrated a Swin-Transformer with an inverted residual module to extract spatial features from micro-Doppler maps, achieving 99.2% accuracy across seven activities. Kang et al. [91] proposed a spatial–temporal point cloud transformer (ST-PCT), which operates directly on unstructured point clouds without voxelization, obtaining 99.06% accuracy on the MM-Act dataset. It demonstrated that transformers can process irregular radar representations beyond spectrograms. Gu et al. [116] further advanced the design by combining CNN-based spatial encoding with a Swin-Transformer temporal encoder, achieving 97.5% accuracy and a relative gain of up to 14.8% over lightweight baselines. This hybrid approach underscores a recurring trend: CNNs remain effective for local pattern extraction, while transformers excel at modeling global dependencies. Qu et al. [74] addressed HAR in tangential motions exceeding 60°, where micro-Doppler cues weaken. Their SCL-ST network leveraged Swin-Transformer encoders to enhance local–global feature fusion, improving recognition to 98.89% even under subtle motion variations.

Table IX summarizes the performance of transformer-

TABLE IX: Summary of studies utilizing transformer models.

Study	Experimental Data	Performance Details
[57] (2023)	12 subjects, 7 designed activities, totaling 3,279 samples	The proposed IR-ST model combines an inverted residual module to extract spatial feature information from a micro-Doppler map and a Swin-Transformer module that employs a self-attention mechanism to learn timing feature information, achieving recognition accuracy at 99.2%
[91] (2023)	MM-Act dataset: Each person performs the action 0.8-1 m in front of the radar, 21 single persons (including 16 males and 5 females), 9 designed activities in place, a total of 15,627 samples.	ST-PCT proposed featuring a spatial neighbour embedding module, a temporal and spatial attention mechanism, and an optimised attention mechanism, achieving 99.06% accuracy on the MM-Act dataset without voxelisation.
[116] (2024)	Glasgow dataset.	RMPCT-Net used a CNN and a transformer for feature extraction of the time-range map and micro-Doppler map. It has achieved an average accuracy of 97.5%. Additionally, the model displays an average accuracy increase of 1.3%–14.8% compared to traditional lightweight networks
[74] (2025)	The dataset was collected from an indoor laboratory, consisting of five males and three females performing six activities, and a total of 1500 data samples were obtained.	SCL-ST network integrates the feature fusion, ST encoder, and SCL framework, with single-channel micro-Doppler spectrograms and dual-channel interferometric spectrograms as input. Achieve a recognition accuracy of 98.89%.

based HAR systems, indicating consistently high accuracies (95–99%) across diverse radar modalities. While the results confirm the strong modeling capacity of transformers, several insights emerge: (i) transformer–CNN hybrids currently outperform standalone transformer designs by balancing local and global features; (ii) lightweight adaptations (e.g., FML-ViT) remain necessary for edge deployment, given the high memory and compute demands of self-attention; and (iii) transformer robustness has been validated not only on micro-Doppler maps but also on irregular point cloud data, highlighting their versatility.

Overall, transformer models represent the latest evolutionary step in deep learning for FMCW radar HAR. Their ability to unify multi-domain representations makes them promising candidates for real-world deployment, though future work must address challenges in model compression, interpretability, and generalization across domains.

5) Hybrid and Multimodal Fusion models: Hybrid models are designed to address the limitations of CNN-RNN approaches by combining different architectural strengths and exploiting information from diverse data sources. They enhance the ability to capture complex dynamics and improve performance in real-world environments, particularly when labeled data are scarce or the operating conditions are highly variable.

Collecting and labeling radar data for HAR is inherently challenging, which has motivated the adoption of unsupervised and self-supervised strategies. Compared with earlier sections where AEs were mainly used for signal reconstruction, in hybrid settings, they serve as part of a broader pipeline that integrates discriminative models. For instance, Yao et al. [93] developed an unsupervised fall detection system that first trains a 3-D AE to reconstruct sequences of range-Doppler frames. Bottleneck features from the encoder-decoder are then passed into a DNN predictor to determine a fall threshold. This approach eliminates the need for labeled fall samples and achieves an accuracy of 95.54% with a false alarm rate of 1.07% across multiple environments and subjects.

Another direction leverages contrastive learning to improve generalization. While micro-Doppler signatures are powerful for HAR, they are sensitive to aspect-angle variation, which severely affects robustness. Qiu et al. [123] addressed this by employing SimCLR [124], a CNN-based contrastive framework, to learn transformation-invariant micro-Doppler features. The learned embeddings were shown through t-SNE visualization to cluster activities across different orientations, supporting omnidirectional HAR. Other hybrid approaches incorporate CNN modules with Transformers, such as IR-ST [57] and RMPCT-Net [116], which integrate local convolutional features with global self-attention. Although these studies were described in the Transformer subsection, they are better understood here as hybrids since they explicitly fuse CNN and Transformer components. Table X summarizes representative studies on hybrid models.

In addition, radar signals contain rich characteristics that cannot be fully expressed by a single domain of feature representation. Exploring efficient ways to utilize radar features from multiple domains has attracted considerable research interest [101], [105]–[108], [113]. The main pipeline involves extracting features from different domains using separate feature learning networks and then fusing the learned features to make comprehensive final predictions, as illustrated in Fig. 10. Applying multimodal fusion offers two main advantages: completing missing information from a single domain and detecting relationships between two or more different domains.

In [106], latent representations of range-Doppler, range-time, and micro-Doppler maps were extracted using three pre-trained AE networks, which were subsequently fed into an LSTM network for sequence modeling and prediction. Similar approaches that synergistically exploit time-range-Doppler features have also been observed in [101], [113], where different feature maps were modeled using multi-branch CNN-based architectures. These approaches have demonstrated the superiority of multi-domain methods compared to single-domain methods across multiple datasets, particularly in complex situations such as random positions or occluded environments.

In real-world situations, such as predicting activities from arbitrary trajectories, angle information from MIMO radar is also considered to provide spatial cues for the HAR system [46], [69], [107], [108]. [46] investigated a fall detection system utilizing range-Doppler and range-angle maps. Vari-

TABLE X: Summary of studies utilizing hybrid models.

Study	Experimental Data	Performance Details
[38] (2020)	mmFall dataset [38]	Used point cloud data and a hybrid variational RNN-Autoencoder to identify anomalies in body motion, achieving 98% detection accuracy with minimal false alarms.
[93] (2023)	Radar is 1.6 m from the ground with a scan range of 0–5 m, 21 subjects, 10 different real environments, 52 kinds of daily non-fall actions and 12 kinds of fall actions, totaling 28,440 samples	Autoencoder proposed to extract range–velocity–time features+DNN proposed predictor. The model is trained solely on unlabeled non-fall samples, achieving an accuracy of 95.54%, a false alarm rate of 1.07%.
[57] (2023)	12 subjects, 7 designed activities, totaling 3,279 samples	The proposed IR-ST model combines an inverted residual module to extract spatial feature information from a micro-Doppler map and a Swin-Transformer module that employs a self-attention mechanism to learn timing feature information, achieving recognition accuracy at 99.2%
[123] (2024)	The radar is set at point O_1, the target is located 4 m from the radar, which is set at a height of 0.8 m with a 0° pitch angle, and at point O_2, the target is 8 m away from the radar, with the radar heights set at 0.5 m and 1.5 m, 6 designed activities, a total of 1,440 samples.	A stacked dual-path feature fusion network (SDP-Net) is designed for omnidirectional human motion recognition by fusing micro-Doppler features in the target motion direction. Its lightweight variants achieve superior performance in activity recognition.
[116] (2024)	Glasgow dataset.	RMPCT-Net used a CNN and a transformer for feature extraction of the time-range map and micro-Doppler map. It has achieved an average accuracy of 97.5%. Additionally, the model displays an average accuracy increase of 1.3%–14.8% compared to traditional lightweight networks

ous feature fusion methods for improving overall accuracy were examined, including prediction weighting, voting, and ensemble learning. To address the issue of varying aspect angles, Zhao et al. [69] introduced two point transformation networks [127] to convert the multi-angle point cloud data into a uni-angle format and classify the point cloud into groups of activities. The output from the point module is then combined with micro-Doppler spectrograms to serve as input data representations for the AlexNet HAR classifier. Refer to Table XI for additional details of multimodal fusion models.

6) Lightweight-Oriented and Efficient models: Deep learning is inherently data-driven, yet the available datasets for HAR with FMCW radar remain limited, and real-world deployments often face resource constraints. A radar spectrogram can be considered an RGB image. When using images as input, a natural approach is to apply models designed for image processing, which have achieved good results in HAR-based radar tasks. One challenge in radar-based HAR is the small number of samples in datasets, which makes it difficult for the model to converge and may cause it to get stuck in a local minimum [128], resulting in performance that may not match that of image processing tasks. In addition, radar-based HAR data is often collected in limited, controlled environments with a restricted number of participants. Consequently, researchers have proposed three complementary strategies to enhance both efficiency and generalization: (i) simulation-based data generation to mitigate data scarcity, (ii) domain adaptation and transfer learning to address distribution mismatch, and (iii) lightweight architectures tailored for low-latency deployment. Key findings from studies related to these techniques are presented in Table XII.

In a different approach, [48] designed a human motion synthesis tool that simulates raw radar data, completely eliminating the data collection process. The authors placed 21 virtual markers on a 3D humanoid character in animation software and captured time-variant 3D trajectories of the markers. These 3D marker trajectories are utilized by a geometrical 3D indoor channel model to simulate diverse human activities. The HAR system trained on this fully simulated dataset yields a remarkable accuracy of 98.4% on real unseen data. Additionally, [50] extended this simulation-based approach to a 2x2 MIMO radar system for direction-independent HAR.

Since indoor radar signals are highly environment-dependent, several studies have investigated effective ways to adapt radar-based HAR systems to unseen domains [40], [49], [58], [65], [66], [72]. Li et al. [40], [49] proposed adversarial training frameworks to learn domain-invariant features and then utilized transfer learning to leverage prior knowledge from the source domain, allowing the system to generalize to the target domain with limited labeled data. To effectively learn category-discriminative features, [65], [66] employed metric-learning frameworks to enable the system to discriminate features in a self-supervised manner.

7) Emerging Few-shot and Edge transformers: Data scarcity remains a central bottleneck in FMCW radar-based HAR, particularly when systems must operate in diverse real-world environments where collecting large, labeled datasets is impractical and costly. Few-shot learning has therefore emerged as a key paradigm, enabling deep models to generalize from only a handful of annotated samples while reducing annotation effort and improving adaptability. Chen et al. [98] tackled the challenge of limited labelled radar data in real-world scenarios by DSMFT-Net, which combines clustering and contrastive learning to develop robust, domain-invariant feature representations for few-shot HAR, achieving 93.3% under 5-shot and 96.5% under 10-shot settings. Li et al. [49] introduced supervised few-shot adversarial domain adaptation (FS-ADA), enabling effective model training with limited data from new scenarios by learning a common feature space and optimizing through multitask generative adversarial training, with 20 samples per class achieving 91.6% accuracy. Complementary strategies, such as wavelet-based feature extraction combined with micro-Doppler signatures [66], have further demonstrated that few-shot approaches can maintain near

TABLE XI: Summary of studies utilizing multimodal fusion models.

Study	Experimental Data	Performance Details
[106] (2022)	Including 4 datasets, the radar mounted in a wall position at a height of 1.5 m, 1 subject, 4 designed activities, totaling 83,536 samples	Multi-View CNN-LSTM architecture fuses multiple views of time-range Doppler for human activity recognition. Using CNN to extract features from various projections of the radar data cube, followed by LSTM networks for sequence classification, achieving a Macro $F1$ score of 74.7%
[46] (2022)	The radar was fixed at a height of 1.6 m above the ground, 21 subjects, more than 50 different types of non-fall actions, and 12 fall actions, totaling 12,720 samples	Utilizing CNN addresses challenges in fall detection in real-world scenarios by employing ensemble learning to fuse features from multiple signal maps, achieving a high recall of 0.983 and precision of 0.975 in diverse environments.
[69] (2022)	Experiments were conducted in an office-like room with tables, chairs, and cabinets. The radar set as 0.75 m in height, 8 subjects, 6 designed activities, totaling 2,239 samples	T-Net and PointNet using six intrinsic features—range, azimuth, elevation, Doppler, received power, and time—by combining point clouds and spectrograms. The orientation classification module provided very accurate predictions with an accuracy of 97.5%, the Point Cloud (PC) classification module achieved an average accuracy of 97.9%, while the spectrogram classification module had a lower accuracy of 83.8%.
[108] (2023)	The radar is set as 2.5 m in height and slopes down 30°, 4 experimenters, 5 designed activities, totaling 24,000 samples	Focuses on a fall detection method using parallel 2DCNN-CBAM with radar multidomain representations to reduce false alarms, achieving the fall detection average accuracy of 96.7%.
[113] (2023)	RadHAR dataset [125]	Combining a temporal feature extraction stream (TFES) and a spatial feature extraction stream (SFES), utilizing a coordinates-based spatial attention mechanism (CSAM) to enhance spatial feature extraction. Achieving an accuracy of 97.10% on the RadHAR dataset.
[99] (2023)	Human Activity Data with a 5.8GHz FMCW Radar dataset [99]	Combine sparse theory and PointNet network with both operations in the time-Doppler (TD) and range-Doppler (RD) domains. The TDSP method and DRDSP method achieve an average accuracy of 95.0% and 98.0%
[107] (2023)	The indoor experiments are set in an empty room with radar set as 1.2 m in height, 5 designed activities, totaling 2,576 samples with one subject and 1,920 samples of two-person activity data	The Doppler, range, azimuth, and elevation features of the point clouds as inputs were used to train a four-channel convolutional neural network (CNN) classification model with channel attention, achieving accuracy results of 96.09% for the indoor-MPAR task.
[105] (2024)	Glasgow dataset	A novel lightweight multidomain multilevel fused patch-based learning model (convolutional mixer) utilizing an early fusion of RGB color domain information from 2-D radar representations, achieving an accuracy of 91.05%
[101] (2024)	Millimeter wave radar data of people walking dataset [126]	POMEN model utilizes 3D convolutions in the earlier stages to capture the information embedded in the sparse 3D TRD representation, and adds the 2D features corresponding to the range-time signature, range-Doppler map, and time-Doppler signature (spectrogram), respectively. Achieving an accuracy of 89.45% with a 3D TRD cube.
[87] (2025)	The radar was mounted on a stand at a height of approximately 2.8 m above the ground in a 6 m x 5 m indoor area, with 10 participants performing six different types of activities from 12 different locations, totaling 1,640 samples of point cloud data.	The point cloud input used for the multimodal feature fusion network achieved an average recognition accuracy of 98.77%, which is superior in computational efficiency compared to other networks such as 3D-GCN, DGCNN, PointNet, and PointNet++.
[117] (2025)	Positioning the radar directly above the tester at approximately 3 m from the ground, the dataset comprised a total of 12 testers performing seven actions sequentially, resulting in 10,005 samples.	MFECNet used both range-time maps and Doppler-time maps as inputs to recognize fused features, achieving an overall accuracy of 99.67% on its self-built dataset. Additionally, improvement of 1.62%–3.62% compared to other models using the same Glasgow dataset.
[112] (2025)	The dataset was collected from 13 different environments and included 16 volunteers, consisting of 13 males and 3 females, who performed four types of falls and twenty types of non-fall motions, totaling 43,577 samples.	mm-Fall used angle-range maps to identify and locate multiple moving targets. It can adapt to new domains and effectively distinguish falls from non-fall motions, with an average recall of 96.9% and a precision of 99.6% in fall detection.

state-of-the-art accuracy (98.33%) with as few as 30 training samples. Beyond data efficiency, recent work has begun to explore edge-oriented transformer architectures, aiming to reduce computational overhead while retaining the attention-based modeling capacity of large networks. Ding et al. [70] focused on efficiency, introducing FML-ViT, a lightweight architecture using linear self-attention, which reached 95.1% accuracy while reducing complexity compared to conventional ViTs. Although still nascent, these approaches are critical for enabling real-time HAR on resource-constrained edge devices.

Overall, as summarized in Table XIII, few-shot learning frameworks consistently report accuracies above 90%, even under highly limited training conditions. This demonstrates their potential to close the gap between simulated and real-world data, while also highlighting the need for future integration with lightweight edge transformers to ensure practical deployment in latency- and energy-constrained environments.

D. What Public FMCW Radar Datasets Exist To Support Reproducible Research?

Collecting data using radar requires considerable time and effort, leading to a limited number of publicly available datasets in HAR research. While previous publications [25], [29]–[31] have mentioned some publicly available datasets for various types of radar, none have comprehensively and thoroughly compiled the publicly available datasets specifically for FMCW radar. Therefore, in this section, we will undertake the work outlined above by focusing on important

TABLE XII: Summary of studies utilizing lightweight-oriented and efficient models.

Study	Experimental Data	Performance Details
[40] (2021)	Glasgow dataset	Joint domain and semantic transfer learning (JDS-TL) utilizes a sparsely labeled dataset to train the HAR model, achieving an average accuracy of 87.6% in recognizing six human activities with only 10% of the instances labeled in the training dataset.
[65] (2021)	Glasgow dataset	The transfer learning mechanism enables adaptation to new environments without source data, utilizing a self-supervised labeling strategy combining neighbor-aggregating and clustering methods to enhance label reliability for target-specific feature extraction, achieving a mean accuracy of over 93%.
[48] (2023)	The simulated radar dataset consisted of a total of 1,200 samples of 5 activities	The simulation framework generates realistic radar signatures, thus minimizing the need for large experimental datasets. The CNN model was used, and an average classification accuracy of 94%.
[50] (2024)	The simulated radar dataset consisted of a total of 5,652 samples of 5 activities	Introduces a simulation-based approach to address the scarcity of radar data for multiple input multiple output radar (MIMO), by synthesizing human motion and generating spatial trajectories to create realistic MIMO radar signatures. The simulation method on MoCap achieves a classification accuracy of 97.83%.
[58] (2024)	Experiments were conducted in a 10 × 6 m area, setup two different frequencies of radar for dataset acquisition, 18 subjects, 6 designed activities, and a total dataset of 2,160 samples.	The Heterogeneous Multi-source Transfer Learning (HMTL) method is based on the Categorical Probability to obtain Hash Coding (CPHC) algorithm for multivariate data selection in the radar micro-Doppler signature domain, achieving a classification accuracy of 99.2%.
[104] (2024)	Proposed Dataset: The dataset contains ten scenarios and 21 subjects performing 52 kinds of nonfall activities and 12 kinds of fall activities, a total of 28,440 samples. Public Dataset: the Glasgow dataset.	The proposed method utilized a ResNet feature extractor to extract domain-invariant features and a multilayer perceptron as a classifier for fall detection to enhance generalization across different users and environments, achieving high F1 scores of 97.47% and 99.49% on respective datasets.
[114] (2024)	The heights of the radar from the ground are 2.0 m and 1.8 m in the five different scenes. A total of 40 subjects performed 28 activities divided into two categories: sudden and soft falls.	Lightweight deep learning network used Range-time spectrograms for real-time fall detection, achieving 99.63% accuracy in unseen subjects' fall detection.
[75] (2024)	The radar is mounted on a bracket at a height of 2.8m above the ground in a 5m × 5 m indoor area, with 10 targets performing six distinct daily activity types, and a total dataset of 3,600 samples.	PDHE-Net consists of a lightweight deep learning feature extraction module, a parameter estimation module, and a classification module that extracts deep features from the time-Doppler maps input. The result achieved a recognition accuracy of 96.67% on the dataset.
[72] (2025)	Experiments were conducted in an office, 6 m x 10 m x 3.2 m, and a hallway. The radar sensor is mounted on a wooden pole approximately 1.2 m above the ground, with 6 designed activities. 11 subjects were instructed to perform the actions from various positions and directions (e.g., 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), and a total dataset of 1,894 samples.	SST-DCN is a lightweight dilated convolutional network that extracts spatial-aware, multi-scale spectro-temporal information through multi-channels, achieving a classification accuracy of 90.72%. SST-DCN's accuracy is higher than AlexNet, VGG11, ResNet18, MobileNet V2, and ShuffleNet V2.
[71] (2025)	CI4R dataset.	RadSpecFusion model used a novel attention-based fusion architecture for multi-radar human activity recognition, achieving 99.21% accuracy on the validation set across 11 distinct human activity classes using the CI4R dataset. The confusion matrix revealed minimal misclassifications (an error rate of only 0.8%), occurring primarily between kinematically similar activities.
[77] (2025)	CI4R dataset.	RSA architecture employing adversarial training of the gradient reversal layer and the auxiliary classifier generative adversarial network with Micro-Doppler spectrogram input from radar datasets from three devices. In a single-domain scenario, accuracy improves by at least 1.5% over the baseline. Multidomain scenario, achieving approximately 97% accuracy with three domains.

details such as the name of the dataset, the creation year, the sensor hardware model/vendor, the data format, the FMCW parameters, the antenna layout, the number of subjects, the number of activities, the name's activities, and the collection environment, which can be seen in Table XIV. Fifteen publicly available HAR datasets were identified between 2018 and 2025. Most datasets are provided in raw data format, with additional formats including micro-Doppler spectrogram images [62], [66], [119] and point clouds [38], [66], [125], which enable advanced learning models to extract spatial and temporal features more effectively. Depending on the type of sensor, these datasets use FMCW radars with different center frequencies, such as 5.8 GHz, 24 GHz, 60 GHz, and 77 GHz, where 60 GHz and 77 GHz are the most prevalent due to their fine range resolution and cost availability. Nearly all datasets adopt a single-sensor setup, limiting spatial diversity, with only a few employing two sensors or multi-environment configurations [111], [118], which are essential for testing model robustness in real-world settings. The number of human subjects varies widely, from as few as 1 to as many as 99 subjects [118], with only five datasets exceeding 10 subjects, raising concerns about model generalizability across different body types and motion styles. The annotated activities largely consist of daily actions such as walking, sitting, falling, squatting, and bending. However, some datasets emphasize more complex behaviors such as entering or leaving rooms [109], seizure-like movements [129], or multi-frame fall events [94], offering valuable scenarios for healthcare applications. Despite

TABLE XIII: Summary of studies utilizing emerging few-shot and edge transformers models.

Study	Experimental Data	Performance Details
[49] (2021)	Glasgow dataset	Proposed supervised few-shot adversarial domain adaptation (FS-ADA) enabling effective model training with limited data from new scenarios by learning a common feature space and optimizing through multitask generative adversarial training. Achieves an accuracy of 91.6% with 20 samples per class.
[66] (2023)	Human activity recognition with FMCW radar using few-shot learning datasets	Use wavelet transforms, background frame differences, and micro-Doppler signatures for effective feature extraction, which was used in few-shot learning, achieving 98.33% accuracy with only 30 training samples.
[70] (2024)	Self-Datasets: The dataset involves 15 volunteers performing 6 distinct activities, resulting in a total dataset size of 4,500. Open-Source Datasets: the CI4R dataset.	FML-Vit is a lightweight image transformation algorithm with layer-wise linear self-attention applied to micro-Doppler input. It achieves 95.1% accuracy in activity recognition, which is higher than ten other algorithms evaluated on two different datasets.
[98] (2025)	The radar was positioned in an area of 2 m in width and 4.1 m in length, with eight volunteers (5 males and three females) performing six types of activities.	DSMFT-Net is a dual-input network that combines preprocessed simulated radar Doppler spectrograms and measured radar Doppler spectrograms. The core challenge addressed is the scarcity of labeled radar samples and stark disparities in data distribution between simulated and measured activity domains, with an average accuracy of 93.3% under a 5-shot setting and 96.5% under a 10-shot setting across six human activity recognition tasks.

their utility, most datasets are captured in controlled indoor environments with subjects performing tasks in front of the radar, limiting the domain variability and thus the transferability of trained models. Notably, only one dataset includes data from eight distinct environments [118], underscoring a lack of environmental diversity in current public datasets. In the articles analyzed, while deep learning models trained on these datasets exhibit promising results in terms of classification accuracy within the training domain, performance degrades significantly when applied to unseen users, sensor positions, or dynamic environments. Due to the scarcity of large datasets, models have difficulty extracting domain-invariant features. Therefore, future datasets should increase the number of samples, participants, and environmental diversity to promote more generalization and robustness of HAR models.

TABLE XIV: Technical specifications of the publicly available datasets.

Dataset	Year	Sensor Hardware Model/Vendor	Data Format	FMCW Parameters	Antenna Layout	No. of Subjects	No. of Activities	Name's Activities	Env
[130]	2018	Industrial Radar Systems GmbH	Raw data	Fc=77GHz, B=1.5GHz, Tc=256μs, Fs=2GHz, Ns=256, Nc=256	SISO	5	1	Walking	2
[118]	2019	Ancortek	Raw data	Fc=5.8GHz, B=400MHz, Tc=1ms, Fs=N/A, Ns=128, Nc=N/A	SISO	99	6	Walking, sitting down, standing up, pick up an object, drink water and fall	8
[109]	2019	INRAS GmbH	Raw data	Fc=77GHz, B=N/A, Tc=256μs, Fs=N/A, Ns=160, Nc=256	SISO	9	6	Entering room, leaving room, sitting down, standing up, clothe, unclothe	1
[125]	2019	IWR1443 Texas Instruments	Point clouds	Fc=77GHz, B=4GHz, Tc=N/A, Fs=N/A, Ns=N/A, Nc=N/A	MIMO	2	5	Walking, Jumping, Jumping Jacks, Squats, Boxing	1

[129]	2019	AWR1642 Texas Instruments	Point clouds	Fc=77GHz, B=3.072GHz, Tc=N/A, Fs=2.5MHz, Ns=128, Nc=256	MIMO	6	6	Other behavior, walking, falling, swinging hand for help, seizure, restless movement	1
[126]	2020	AWR1642 Texas Instruments	Raw data	Fc=60GHz, B=3.6GHz, Tc=160μs, Fs=10MHz, Ns=512, Nc=128	MIMO	29	6	Fast walking, slow walking, low walking with hands inside pockets, walking with a metallic bottle under the jacket, walking with a limp, walking at slow speed with swinging hands	1
[119]	2020	SDR-Kit Ancortek (24GHz radar), IWR1443 Texas Instruments (77GHz radar)	Raw data, micro-Doppler spectro-gram	24GHz radar: Fc=24GHz, B=1.5GHz; 77GHz radar: Fc=77GHz, B=750MHz	SISO	6	11	Walking towards the radar, picking up an object from the ground, sitting on a chair, crawling towards the radar, walking on both toes, scissor gait, walking away from the radar, bending, kneeling, limping with right leg stiff, walking with short steps	1
[38]	2020	AWR1843 Texas Instruments	Point clouds	Fc=77GHz, B=1.92GHz, Tc=N/A, Fs=2MHz, Ns=128, Nc=64	MIMO	2	9	Randomly walking, forward fall, backward fall, left fall, right fall, sitting down on the floor, crouching, bending, jump	1
[111]	2022	xWR14xx, xWR68xx Texas Instruments	Raw data	Fc=77GHz and 60GHz, B=4GHz, Tc=80μs, Fs=5MHz, Ns=93, Nc=128	MIMO	9	14	Walk to room, fall on the floor, stand up from the floor, walk to chair, sit down on chair, stand up from chair, walk to bed, sit down on bed, stand up from bed, get in bed, lie in bed, roll in bed, sit in bed, get out bed	2
[99]	2022	Nanjing University of Science and Technology and Texas Tech University	Raw data	Fc=5.8GHz, B=320MHz, Tc=3.3ms, Fs=N/A, Ns=N/A, Nc=N/A	MIMO	16	6	Falling, jumping, jogging, walking, stepping, and squatting	2
[66]	2023	IWR1443 Texas Instruments	micro Doppler	Fc=77GHz, B=3.16GHz, Tc=N/A, Fs=N/A, Ns=128, Nc=128	MIMO	7	8	Bow. boxing, hand up, fall down, walk, squat, sit, stand	3
[96]	2023	Position2Go Infineon	Raw data	Fc=24GHz, B=170MHz, Tc=300μs, Fs=N/A, Ns=64, Nc=64	SISO, MIMO	2	3	Handwaving, walking, running	1
[131]	2024	Simhumalator software [132]	Raw data	Fc=24GHz, B=400MHz, Tc=1ms, Fs=N/A, Ns=128, Nc=N/A	SISO	1	11	Walking, body rotation, punching, kicking, grabbing, standing up, sitting down, standing to walk, walking to sit, walking to fall, and falling to walk	1
[94]	2024	IWR6843 Texas Instruments	Raw data	Fc=60.25GHz, B=3.6GHz, Tc=160μs, Fs=6MHz., Ns=128, Nc=128	MIMO	10	5	Bending, falling, sitting, squatting, walking	1

[62]	2025	IWR1843 Texas Instruments	micro Doppler	Fc=77GHz, B=2998.2MHz, Tc=N/A, Fs=10MHz, Ns=128, Nc=255	SISO	10	6	Bending; falling; jumping; punching; jogging; and walking	1

N/A: Not available; Fc: Center freq; B: Bandwidth, Tc: Chirp duration; Fs: Sample freq; Ns: Samples per chirp; Nc: Chirps per frame
SISO: Single-Input Single-Output ; MIMO: Multiple-Input-Multiple-Output

Table XV presents an overview of the publicly available datasets employed in the analyzed literature to train, test, and evaluate deep learning models in FMCW radar-based HAR. Among the 15 identified datasets ranging from 2018 to 2025, the dataset from Fioranelli et al. [118] stands out as the most widely adopted, referenced in 22 articles, reflecting its comprehensive activity set and diverse environmental scenarios. This dataset's size is 1.9 GB, and its raw data format facilitates feature extraction, aligning well with the demands of multi-domain learning frameworks. The mmFall dataset [38], despite its compact size (453.92 MB), is less frequently employed, appearing in only two studies. Several large-scale datasets, such as PARrad [111] (809.6 GB) and CI4R [119] (86.8 GB), offer rich, multi-environment data conducive to improving model robustness and generalization, yet their utilization remains relatively low, indicating barriers possibly related to environment setup and processing complexity. Notably, some foundational datasets like IDRad [130] and the Multiple Patients Behavior Dataset [129] remain underexplored in recent literature, revealing gaps in dataset adoption and potential opportunities for broader evaluation. The predominance of raw data formats, alongside micro-Doppler spectrograms and point cloud representations, supports the extraction of invariant and discriminative features crucial for addressing variability in real-world human activities. To aid further research, direct download links to these datasets are provided, ensuring accessibility and encouraging experimentation with datasets within the community.

IV. Discussion

A. Challenges

Answering the research questions in the previous section confirms that deep learning-based HAR with FMCW radar has advanced significantly between 2018 and 2025. Nonetheless, the field faces persistent obstacles that constrain scalability and real-world deployment, as detailed below.

Limited Dataset Size and Diversity: Most studies rely on small, homogeneous datasets, which undermines model generalizability in unseen environments or with diverse users [37], [41], [48], [50], [52], [62], [95]. Only a few datasets, such as those in [66], [111], [118], [119], [126], include sufficient variation in subject demographics or environmental settings. By contrast, many collections remain restricted to a few class labels and participants [37], [48], [52], [96], [129], [130]. This imbalance reinforces the need for large-scale, publicly available datasets with broader contextual diversity.

Difficulty in Distinguishing Similar Activities: Deep models often struggle to discriminate between activities with similar micro-Doppler spectrograms (e.g., bending to pick up an object versus drinking water, or falling versus fall-like actions) [67], [73], [95], [133]. The subtle spectral differences are challenging to detect without domain-robust feature representations. This limitation underlines the demand for better feature extraction and model architecture to detect slight differences between similar activities. Recent efforts leveraging multi-modal fusion or attention-based feature learning [46], [66], [108], [128] represent initial steps but remain insufficient for high-fidelity recognition in real-world conditions.

Environmental Interference and Signal Noise: Radar performance deteriorates in cluttered or uncontrolled environments due to multipath propagation and reflective noise [54], [60], [103], [134]. While denoising methods such as adaptive filtering [39], [56]–[58], [60], [115] offer partial solutions, maintaining consistent accuracy across heterogeneous scenarios remains unresolved. Real-time temporal–spatial denoising and domain adaptation strategies are promising but not yet mature.

Computational Complexity and Real-Time Constraints: State-of-the-art networks, including CNN–RNN hybrids and 3D-CNNs, often impose prohibitive computational costs that hinder edge deployment [48], [52], [60], [66], [67], [99], [101], [106], [117]. Although lightweight alternatives and efficient model compression techniques [70], [72], [78], [98] have been proposed, balancing latency, energy consumption, and accuracy remains a critical challenge.

B. Future Research

Despite encouraging progress, the limitations identified above restrict large-scale translation of FMCW radar HAR into practical systems. Future research must therefore concentrate on addressing data scarcity, recognition robustness, noise resilience, and algorithmic efficiency. We outline four strategic directions:

Data Augmentation and Expanding Datasets: Increasing the diversity of subjects, activities, and environmental conditions is vital to cultivate generalizable models [50], [51], [66], [104], [106]. Comprehensive datasets such as Radar signatures of human activities [118], Human activity recognition with FMCW radar using few-shot learning Datasets [66], PARrad [111], and CI4R [119] demonstrate the utility of broader environmental scenarios and multi-user configurations. Future initiatives should prioritize scale, accessibility, and standardized benchmarks.

Enhancing Recognition Accuracy for Similar Actions: Concerning activities with closely aligned micro-Doppler signatures, such as actual falls versus fall-like motions, remains

TABLE XV: Summary of publicly available FMCW radar datasets for HAR.

Dataset	Size	No. of Samples	Articles Utilized Dataset	Download Link	Highlights
IDRad [130]	15.7 GB	1,490	[130]	IDRad	Built a solution to Automatic indoor person identification using gait features.
Radar signatures of human activities [118]	1.9 GB	1,754	[37], [40], [43], [45], [49], [51], [53], [58], [59], [62], [65], [67], [68], [80], [83], [93], [101], [102], [104], [105], [113], [116]	Glasgow	Provide a dataset for the development and evaluation of feature extraction and classification algorithms for human activities.
HARrad [109]	27.4 GB	1,505	[109]	HARrad	Develop for the HAR.
RadHAR [125]	881 MB	15,635	[89], [95], [113]	RadHAR	Make use of sparse and heterogeneous point clouds for accurate HAR.
Multiple Patients Behavior Dataset [129]	105.2 MB	105,423	[129]	Behavior	Assist in simultaneously recognizing multiple patients' behaviors by detecting the scattered point cloud.
Millimeter wave radar data of people walking [126]	27.57 GB	1,600	[101]	Part 1: Record1 Part 2: Record2	Evaluate different ways of walking of people for human gait recognition.
CI4R [119]	86.8 GB	9,900	[62], [70], [71], [77], [83]	CI4R	Monitor activities of daily living and recognize ambulatory gait.
mmFall [38]	453.92 MB	N/A	[38], [83], [93]	mmFall	Identify a human fall via point cloud data.
PARrad [111]	809,6 GB	21,569	[111]	Parrad	Resolve an issue for automatic HAR in hospital rooms.
Human Activity Data with a 5.8 GHz FMCW Radar [99]	2.63 GB	1,440	[76], [99]	DataHAR	Establish a radar-based system for monitoring human activities.
Human activity recognition with FMCW radar using few-shot learning Datasets [66]	6.37 MB	2,520	[66]	Fewshot	Implement the few-shot learning algorithm for HAR.
Multiple Individual Human Activity Recognition Using Microwave FMCW Radar [96]	177 MB	3,455	[96]	Multiple	Recognizing human activities in multi-subjective environments.
FMCW radar dataset [131]	10.49 GB	19,800	[131]	FMCWdataset	Enhance human activity classification utilising micro-Doppler signatures with the effects of noise.
FallDetection-base-on-FMCW [94]	1.7 GB	4,620	[94]	FallDetection	Designs a continuous fall detection system using multiframe input.
I+ Lab RadSet v1 [62]	29.46 MB	4,120	[62], [83]	RadSet	Provides a comprehensive foundation to evaluate networks that can detect falls and daily activities.

N/A: Not available

a nuanced challenge. Recent works [66], [66], [68], [72], [93], [103], [104] have begun exploring advanced feature extraction and few-shot learning techniques to amplify subtle discriminative cues, thereby significantly elevating classification fidelity in complex contexts.

Noise Reduction: Robust noise suppression methods tailored to minimize environmental interference. Adaptive filtering and signal enhancement methods [39], [56]–[58], [60], [92], [115] provide early solutions, but further progress will depend on combining denoising with cross-domain adaptation to preserve invariance across different radar setups.

Algorithm Enhancement: Advancing algorithmic efficiency by reducing model complexity without sacrificing recognition power represents a vital step toward practical deployment. Lightweight architectures capable of efficiently parsing spectrally similar activities, as investigated in recent studies [40], [57], [64], [70], [72], [98], [114], [116], [123], will enable minimizing the parameters of deep learning models, saving resources, and be adaptable to diverse feature domains.

V. Conclusion

Human activity recognition (HAR) systems have been regarded as a critical research frontier, especially within healthcare domains such as elderly monitoring in clinical and independent living settings. This paper has presented a systematic PRISMA-based survey of deep learning methods for HAR using FMCW radar, guided by four research questions spanning state-of-the-art approaches, signal processing techniques, model evolution, and dataset availability. Focusing on FMCW radar's capacity to simultaneously capture range, Doppler, and angular information demonstrates its effectiveness as a sensing modality, with micro-Doppler signatures standing out as the most informative descriptors of human activity. Our synthesis highlights the progression from CNNs and RNNs to autoencoders, hybrid frameworks, and multi-domain strategies, reflecting a clear trend toward improved recognition performance and robustness. Furthermore, the review catalogues publicly available FMCW radar datasets, enabling standardized benchmarking and comparative evaluations. Although considerable progress has been achieved, significant challenges remain, including the need for scalable datasets, distinguishing highly similar activities, meeting real-time computational constraints, and ensuring robustness in cluttered environments. Looking ahead, this survey underscores the importance of creating large-scale open datasets, advancing lightweight deep models for edge deployment, and integrating robust noise suppression strategies. Addressing these challenges will accelerate the transition of FMCW radar-based HAR from controlled experiments to reliable real-world applications.

REFERENCES

[1] WHO, "Ageing and health." https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, 2024.

[2] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, "A review on recent advances in doppler radar sensors for noncontact healthcare monitoring," *IEEE Transactions on microwave theory and techniques*, vol. 61, no. 5, pp. 2046–2060, 2013.

[3] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Systems with Applications*, vol. 137, pp. 167–190, 2019.

[4] S. K. Gharghan and H. A. Hashim, "A comprehensive review of elderly fall detection using wireless communication and artificial intelligence techniques," *Measurement*, p. 114186, 2024.

[5] Z. Wu, X. Chen, Y. Lin, J. Wen, and Y. Chen, "A smart home energy management system based on human activity recognition and deep reinforcement learning," *Energy and Buildings*, vol. 325, p. 114951, 2024.

[6] S. Kalimuthu, T. Perumal, R. Yaakob, E. Marlisah, and L. Babangida, "Human activity recognition based on smart home environment and their applications, challenges," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 815–819, IEEE, 2021.

[7] S. S. Yadav, R. Agarwal, K. Bharath, S. Rao, and C. S. Thakur, "Tinyradar: Mmwave radar based human activity classification for edge computing," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2414–2417, IEEE, 2022.

[8] A. Ezzouhri, Z. Charouh, M. Ghogho, and Z. Guennoun, "Robust deep learning-based driver distraction detection and classification," *IEEE Access*, vol. 9, pp. 168080–168092, 2021.

[9] A. Sen, A. Mandal, P. Karmakar, A. Das, and S. Chakraborty, "Passive monitoring of dangerous driving behaviors using mmwave radar," *Pervasive and Mobile Computing*, vol. 103, p. 101949, 2024.

[10] R. Salakapuri, N. K. Navuri, T. Vobbilineni, G. Ravi, K. Karmakonda, and K. A. Vardhan, "Integrated deep learning framework for driver distraction detection and real-time road object recognition in advanced driver assistance systems," *Scientific Reports*, vol. 15, no. 1, p. 25125, 2025.

[11] T.-a. Tran, T. Ruppert, G. Eigner, and J. Abonyi, "Assessing human worker performance by pattern mining of kinect sensor skeleton data," *Journal of Manufacturing Systems*, vol. 70, pp. 538–556, 2023.

[12] S. Mekruksavanich and A. Jitpattanakul, "Automatic recognition of construction worker activities using deep learning approaches and wearable inertial sensors.," *Intelligent Automation & Soft Computing*, vol. 36, no. 2, 2023.

[13] E. Miller, Z. Li, H. Mentis, A. Park, T. Zhu, and N. Banerjee, "Radsense: Enabling one hand and no hands interaction for sterile manipulation of medical images using doppler radar," *Smart Health*, vol. 15, p. 100089, 2020.

[14] H. Kivijärvi and K. Pärnänen, "Instrumental usability and effective user experience: Interwoven drivers and outcomes of human-computer interaction," *International Journal of Human–Computer Interaction*, vol. 39, no. 1, pp. 34–51, 2023.

[15] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *The Visual Computer*, vol. 29, pp. 983–1009, 2013.

[16] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 2259–2322, 2021.

[17] A. Shenoy and N. Thillaiarasu, "A survey on different computer vision based human activity recognition for surveillance applications," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1372–1376, IEEE, 2022.

[18] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[19] L. Gao, G. Zhang, B. Yu, Z. Qiao, and J. Wang, "Wearable human motion posture capture and medical health monitoring based on wireless sensor networks," *Measurement*, vol. 166, p. 108252, 2020.

[20] N. T. T. Hong, G. L. Nguyen, N. Q. Huy, D.-N. Tran, D.-T. Tran, *et al.*, "A low-cost real-time iot human activity recognition system based on wearable sensor and the supervised learning algorithms," *Measurement*, vol. 218, p. 113231, 2023.

[21] L. Fan, Z. Wang, and H. Wang, "Human activity recognition model based on decision tree," in *2013 International Conference on Advanced Cloud and Big Data*, pp. 64–68, IEEE, 2013.

[22] M. H. Kabir, M. R. Hoque, K. Thapa, and S.-H. Yang, "Two-layer hidden markov model for human activity recognition in home environments," *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, p. 4560365, 2016.

[23] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE transactions on geoscience and remote sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.

[24] C. Dewi and R.-C. Chen, "Human activity recognition based on evolution of features selection and random forest," in *2019 IEEE international conference on systems, man and cybernetics (SMC)*, pp. 2496–2501, IEEE, 2019.

[25] M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects," *Computers in biology and medicine*, vol. 149, p. 106060, 2022.

[26] S. Hu, S. Cao, N. Toosizadeh, J. Barton, M. G. Hector, and M. J. Fain, "A survey on radar-based fall detection," *arXiv preprint arXiv:2312.04037*, 2023.

[27] R. C. Tewari, A. Routray, and J. Maiti, "State-of-the-art radar technology for remote human fall detection: a systematic review of techniques, trends, and challenges," *Multimedia Tools and Applications*, vol. 83, no. 29, pp. 73717–73775, 2024.

[28] J. Zhang, R. Xi, Y. He, Y. Sun, X. Guo, W. Wang, X. Na, Y. Liu, Z. Shi, and T. Gu, "A survey of mmwave-based human sensing: Technology, platforms and applications," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2052–2087, 2023.

[29] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "A survey on radar-based continuous human activity recognition," *IEEE Journal of Microwaves*, vol. 3, no. 3, pp. 938–950, 2023.

[30] S. Ahmed and S. H. Cho, "Machine learning for healthcare radars: Recent progresses in human vital sign measurement and activity recognition," *IEEE Communications Surveys & Tutorials*, 2023.

[31] A. Nocera, L. Senigagliesi, M. Raimondi, G. Ciattaglia, and E. Gambi, "Machine learning in radar-based physiological signals sensing: a scoping review of the models, datasets and metrics," *IEEE Access*, 2024.

[32] P. Miazek, A. Żmudzińska, P. Karczmarek, and A. Kiersztyn, "Human behavior analysis using radar data. a survey," *IEEE Access*, 2024.

[33] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *bmj*, vol. 372, 2021.

[34] I. Zupic and T. Čater, "Bibliometric methods in management and organization," *Organizational research methods*, vol. 18, no. 3, pp. 429–472, 2015.

[35] S. Elo and H. Kyngäs, "The qualitative content analysis process," *Journal of advanced nursing*, vol. 62, no. 1, pp. 107–115, 2008.

[36] N. Van Eck and L. Waltman, "Manual for vosviewer version 1.6. 20 (2023)."

[37] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from fmcw radar with bi-lstm networks," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13607–13619, 2020.

[38] F. Jin, A. Sengupta, and S. Cao, "mmfall: Fall detection using 4-d mmwave radar and a hybrid variational rnn autoencoder," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 2, pp. 1245–1257, 2020.

[39] 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 Systems Journal*, vol. 16, no. 2, pp. 3036–3047, 2022.

[40] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.

[41] Y. Kim, I. Alnujaim, and D. Oh, "Human activity classification based on point clouds measured by millimeter wave mimo radar with deep recurrent neural networks," *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13522–13529, 2021.

[42] A. Bhattacharya and R. Vaughan, "Deep learning radar design for breathing and fall detection," *IEEE sensors journal*, vol. 20, no. 9, pp. 5072–5085, 2020.

[43] W. Ding, X. Guo, and G. Wang, "Radar-based human activity recognition using hybrid neural network model with multidomain fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 2889–2898, 2021.

[44] B. Wang, L. Guo, H. Zhang, and Y.-X. Guo, "A millimetre-wave radar-based fall detection method using line kernel convolutional neural network," *IEEE sensors journal*, vol. 20, no. 22, pp. 13364–13370, 2020.

[45] F. J. Abdu, Y. Zhang, and Z. Deng, "Activity classification based on feature fusion of fmcw radar human motion micro-doppler signatures," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 8648–8662, 2022.

[46] Y. Yao, C. Liu, H. Zhang, B. Yan, P. Jian, P. Wang, L. Du, X. Chen, B. Han, and Z. Fang, "Fall detection system using millimeter-wave radar based on neural network and information fusion," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21038–21050, 2022.

[47] Y. Yao, W. Liu, G. Zhang, and W. Hu, "Radar-based human activity recognition using hyperdimensional computing," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 3, pp. 1605–1619, 2021.

[48] S. Waqar and M. Pätzold, "A simulation-based framework for the design of human activity recognition systems using radar sensors," *IEEE Internet of Things Journal*, vol. 11, no. 8, pp. 14494–14507, 2023.

[49] X. Li, Y. He, J. A. Zhang, and X. Jing, "Supervised domain adaptation for few-shot radar-based human activity recognition," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25880–25890, 2021.

[50] S. Waqar, M. Muaaz, S. Sigg, and M. Pätzold, "A paradigm shift from an experimental-based to a simulation-based framework using motion-capture driven mimo radar data synthesis," *IEEE Sensors Journal*, 2024.

[51] Y. Zhou, X. Yu, M. López-Benítez, L. Yu, and Y. Yue, "Corruption robustness analysis of radar micro-doppler classification for human activity recognition," *IEEE Transactions on Radar Systems*, 2024.

[52] C. Li, X. Wang, J. Shi, H. Wang, and L. Wan, "Residual neural network driven human activity recognition by exploiting fmcw radar," *IEEE Access*, vol. 11, pp. 111875–111887, 2023.

[53] L. Qu, Y. Wang, T. Yang, and Y. Sun, "Human activity recognition based on wrgan-gp-synthesized micro-doppler spectrograms," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 8960–8973, 2022.

[54] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 3, pp. 2994–3008, 2022.

[55] X. Yao, X. Shi, and F. Zhou, "Human activities classification based on complex-value convolutional neural network," *IEEE Sensors Journal*, vol. 20, no. 13, pp. 7169–7180, 2020.

[56] B. Wang, Z. Zheng, and Y.-X. Guo, "Millimeter-wave frequency modulated continuous wave radar-based soft fall detection using pattern contour-confined doppler-time maps," *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9824–9831, 2022.

[57] M. Gu, Z. Chen, K. Chen, and H. Pan, "Ir-st: A lightweight transformer network for human fall detection based on fmcw radar," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 25128–25135, 2023.

[58] W. Yin, L.-F. Shi, and Y. Shi, "Indoor human action recognition based on millimeter-wave radar micro-doppler signature," *Measurement*, vol. 235, p. 114939, 2024.

[59] S. A. Shah, A. Tahir, J. Le Kernec, A. Zoha, and F. Fioranelli, "Data portability for activities of daily living and fall detection in different environments using radar micro-doppler," *Neural Computing and Applications*, vol. 34, no. 10, pp. 7933–7953, 2022.

[60] Z. Xu, J. Ding, S. Zhang, Y. Gao, L. Chen, Ž. L. Vasić, M. Cifrek, and Z. D. Chen, "mmcmd: continuous motion detection from visualized radar micro-doppler signatures using visual object detection techniques," *Ieee sensors journal*, vol. 24, no. 3, pp. 3394–3405, 2023.

[61] X. Li, Y. Qiu, Z. Deng, X. Liu, and X. Huang, "Lightweight multi-attention enhanced fusion network for omnidirectional human activity recognition with fmcw radar," *IEEE Internet of Things Journal*, 2024.

[62] W. Ai, H. Xu, J. Li, X. Li, X. Li, Y. Li, S. Li, Z. Xu, and Y. Yu, "Radar-based human activity recognition using time-weighted network based on strip pooling," *IEEE Internet of Things Journal*, 2024.

[63] T. Yang, F. Meng, Q. Xu, and Y.-X. Guo, "Fall feature enhancement and fusion using the stockwell transform with dual mmwave radars," *IEEE Sensors Journal*, vol. 23, no. 2, pp. 1368–1376, 2022.

[64] P. Zheng, A. Zhang, J. Chen, Q. Li, and M. Yang, "Real-time fall recognition using a lightweight convolution neural network based on millimeter-wave radar," *IEEE Sensors Journal*, vol. 24, no. 5, pp. 7185–7195, 2024.

[65] Z. Cao, Z. Li, X. Guo, and G. Wang, "Towards cross-environment human activity recognition based on radar without source data," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11843–11854, 2021.

[66] S. Gong, H. Shi, X. Yan, Y. Fang, A. Paul, Z. Wu, and W. Long, "Human activity recognition with fmcw radar using few-shot learning," *IEEE Sensors Journal*, vol. 23, no. 15, pp. 17815–17824, 2023.

[67] A. Helen Victoria and G. Maragatham, "Activity recognition of fmcw radar human signatures using tower convolutional neural networks," *Wireless Networks*, pp. 1–17, 2021.

[68] L. Liu, S. Wang, C. Song, H. Xu, J. Li, and B. Wang, "Radar-based human motion recognition using semisupervised triple-gan," *IEEE Sensors Journal*, vol. 23, no. 24, pp. 30691–30702, 2023.

[69] Y. Zhao, A. Yarovoy, and F. Fioranelli, "Angle-insensitive human motion and posture recognition based on 4d imaging radar and deep learning classifiers," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 12173–12182, 2022.

[70] M. Ding, G. Dongye, P. Lv, and Y. Ding, "Fml-vit: A lightweight vision transformer algorithm for human activity recognition using fmcw radar," *IEEE Sensors Journal*, 2024.

[71] A. Ibrahim, M. Z. Khan, M. Imran, H. Larijani, Q. H. Abbasi, and M. Usman, "Radspecfusion: Dynamic attention weighting for multi-radar human activity recognition," *Internet of Things*, p. 101682, 2025.

[72] N. C. Hoang, Q. C. Nguyen, M. T. Le, *et al.*, "Advancing robust human activity recognition via informative mmwave radar characteristics and a lightweight spatio-spectro-temporal network," *Measurement*, p. 118056, 2025.

[73] S. Waqar, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed mimo radar system and deep learning," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24916–24929, 2023.

[74] L. Qu, J. Cong, T. Yang, and L. Zhang, "Human tangential activity recognition based on swin transformer and supervised contrastive learning using interferometric radar," *IEEE Sensors Journal*, 2025.

[75] C. Ding, S. Guo, G. Cui, and X. Yang, "A parameter estimation and deep learning hybrid extraction network for multi-directional human activity recognition based on mmwave radar," *IEEE Internet of Things Journal*, 2024.

[76] K. K. Mishra and R. B. Pachori, "Fmcw radar-based human activity recognition based on higher-order synchrosqueezing transform," *IEEE Sensors Journal*, 2025.

[77] Y. Ding, P. Lv, R. Liu, Y. Peng, and M. Ding, "A radar system-agnostic (rsa) learning architecture for human activity recognition," *IEEE Sensors Journal*, 2025.

[78] P. Lei, J. Liang, Z. Guan, J. Wang, and T. Zheng, "Acceleration of fpga based convolutional neural network for human activity classification using millimeter-wave radar," *IEEE Access*, vol. 7, pp. 88917–88926, 2019.

[79] M. Shen, K.-L. Tsui, M. A. Nussbaum, S. Kim, and F. Lure, "An indoor fall monitoring system: Robust, multistatic radar sensing and explainable, feature-resonated deep neural network," *IEEE journal of biomedical and health informatics*, vol. 27, no. 4, pp. 1891–1902, 2023.

[80] D. Wang, C. Zhao, Y. Song, and T. Jin, "Crucial region search and feature discrimination for radar-based human activity recognition," *IEEE Internet of Things Journal*, 2025.

[81] D. Xu, W. Yu, Y. Wang, M. Chen, and Y. Cui, "Human activity detection based on parallel ab-tcn using micro-doppler signatures," *IEEE Sensors Journal*, 2025.

[82] X. Xiong, A. Ren, T. Yuan, A. Zahid, Q. H. Abbasi, and M. A. Imran, "Human motion recognition by micro-doppler features and concatenated cnn-lstm network," *IEEE Sensors Journal*, 2024.

[83] Z. Xu, H. Xu, W. Ai, Y. Yu, Y. Duan, R. Wang, X. Li, and P. Tian, "A radar-based activity recognition approach utilizing an adaptive masking mechanism and gated units," *IEEE Transactions on Aerospace and Electronic Systems*, 2025.

[84] A. Rezaei, A. Mascheroni, M. C. Stevens, R. Argha, M. Papandrea, A. Puiatti, and N. H. Lovell, "Unobtrusive human fall detection system using mmwave radar and data driven methods," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 7968–7976, 2023.

[85] H. Huh, I. Jeong, A. Lee, S. Lee, and Y.-S. Shin, "Leveraging falling acceleration and body part clustering for physics-based human fall detection with millimeter wave radar," *Engineering Applications of Artificial Intelligence*, vol. 159, p. 111500, 2025.

[86] X. Peng, Y. Hu, T. Liu, Y. Wu, T. Saito, and T. Toda, "Stability-enhanced human activity recognition with a compact few-channel mm-wave fmcw radar," *IEEE Transactions on Radar Systems*, 2025.

[87] Y. Wang, W. Kong, M. Zhou, W. Nie, W. He, Q. Zhang, and Y. Pang, "Multi-human activity recognition based on sequential 4d point clouds using frequency-modulated continuous wave radar," *IEEE Transactions on Vehicular Technology*, 2025.

[88] F. Luo, S. Khan, A. Li, Y. Huang, and K. Wu, "Edgeactnet: Edge intelligence-enabled human activity recognition using radar point cloud," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 5479–5493, 2023.

[89] J. Zhu, X. Huang, Z. Deng, and Y. Qiu, "mradhprs: Human pose recognition system from point clouds generated through a millimeter-wave radar," *IEEE Transactions on Aerospace and Electronic Systems*, 2024.

[90] Z. Yu, A. Taha, W. Taylor, A. Zahid, K. Rajab, H. Heidari, M. A. Imran, and Q. H. Abbasi, "A radar-based human activity recognition using a novel 3-d point cloud classifier," *IEEE Sensors Journal*, vol. 22, no. 19, pp. 18218–18227, 2022.

[91] L. Kang, Z. Li, X. Zhao, Z. Zhao, and T. Braun, "St-pct: Spatial–temporal point cloud transformer for sensing activity based on mmwave," *IEEE Internet of Things Journal*, vol. 11, no. 6, pp. 10979–10991, 2023.

[92] W.-L. Hsu, J.-X. Liu, C.-C. Yang, and J.-S. Leu, "A fall detection system based on fmcw radar range-doppler image and bi-lstm deep learning," *IEEE Sensors Journal*, vol. 23, no. 18, pp. 22031–22039, 2023.

[93] Y. Yao, H. Zhang, C. Liu, F. Geng, P. Wang, L. Du, X. Chen, B. Han, T. Yang, and Z. Fang, "Unsupervised-learning-based unobtrusive fall detection using fmcw radar," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 5078–5089, 2023.

[94] L. Yang and W. Ye, "Design of a two-stage continuous fall detection system using multiframe radar range-doppler maps," *IEEE Sensors Journal*, 2024.

[95] C. Gianoglio, A. Mohanna, A. Rizik, L. Moroney, and M. Valle, "On edge human action recognition using radar-based sensing and deep learning," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 4160–4172, 2023.

[96] H. Raeis, M. Kazemi, and S. Shirmohammadi, "Cae-mas: convolutional autoencoder interference cancellation for multiperson activity sensing with fmcw microwave radar," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–10, 2024.

[97] H. Abedi, A. Ansariyan, P. P. Morita, A. Wong, J. Boger, and G. Shaker, "Ai-powered noncontact in-home gait monitoring and activity recognition system based on mm-wave fmcw radar and cloud computing," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9465–9481, 2023.

[98] K. Chen, S. Wang, W. Li, Y. Wang, C. Feng, Y. Zhou, J. Cao, B. Zong, and M. Gu, "A few-shot learning-based dual-input neural network for complex spectrogram recognition system with millimeter-wave radar," *Complex & Intelligent Systems*, vol. 11, no. 6, pp. 1–16, 2025.

[99] C. Ding, L. Zhang, H. Chen, H. Hong, X. Zhu, and F. Fioranelli, "Sparsity-based human activity recognition with pointnet using a portable fmcw radar," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 10024–10037, 2023.

[100] C. Kittiyanpunya, P. Chomdee, A. Boonpoonga, and D. Torrungrueng, "Millimeter-wave radar-based elderly fall detection fed by one-dimensional point cloud and doppler," *IEEE Access*, vol. 11, pp. 76269–76283, 2023.

[101] C. Yao, J. Ren, R. Bai, H. Du, J. Liu, and X. Jiang, "Progressively-orthogonally-mapped efficientnet for action recognition on time-range-doppler signature," *Expert Systems with Applications*, vol. 255, p. 124824, 2024.

[102] W.-Y. Kim and D.-H. Seo, "Radar-based human activity recognition combining range–time–doppler maps and range-distributed-convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[103] Z. Ma, X. Wu, X. Zhang, D. Yao, and W. Deng, "Anti-fixed-interference fall detection based on doppler-time-range maps using millimeter-wave radar," *IEEE Sensors Journal*, 2024.

[104] Y. Yao, H. Zhang, Z. Bai, P. Xia, C. Liu, F. Geng, L. Du, X. Chen, Z. Li, P. Wang, *et al.*, "Unobtrusive high-generalization fall detection: A domain-generalization framework," *IEEE Sensors Journal*, 2024.

[105] A. Dey, S. Rajan, G. Xiao, and J. Lu, "Radar-based human activity recognition using multi-domain multi-level fused patch-based learning," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[106] A. Gorji, A. Bourdoux, S. Pollin, H. Sahli, *et al.*, "Multi-view cnn-lstm architecture for radar-based human activity recognition," *Ieee Access*, vol. 10, pp. 24509–24519, 2022.

[107] Z. Wu, Z. Cao, X. Yu, J. Zhu, C. Song, and Z. Xu, "A novel multiperson activity recognition algorithm based on point clouds measured by millimeter-wave mimo radar," *IEEE Sensors Journal*, vol. 23, no. 17, pp. 19509–19523, 2023.

[108] J. He, Z. Ren, W. Zhang, Y. Jia, S. Guo, and G. Cui, "Fall detection based on parallel 2dcnn-cbam with radar multidomain representations," *IEEE Sensors Journal*, vol. 23, no. 6, pp. 6085–6098, 2023.

[109] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, T. Dhaene, and W. De Neve, "Indoor human activity recognition using high-dimensional sensors and deep neural networks," *Neural Computing and Applications*, vol. 32, pp. 12295–12309, 2020.

[110] T. Stadelmayer, A. Santra, R. Weigel, and F. Lurz, "Data-driven radar processing using a parametric convolutional neural network for human activity classification," *IEEE sensors journal*, vol. 21, no. 17, pp. 19529–19540, 2021.

[111] G. Bhavanasi, L. Werthen-Brabants, T. Dhaene, and I. Couckuyt, "Patient activity recognition using radar sensors and machine learning," *Neural Computing and Applications*, vol. 34, no. 18, pp. 16033–16048, 2022.

[112] C. Zhao, Q. Luo, H. Ding, G. Wang, K. Zhao, Z. Wang, W. Xi, and J. Zhao, "mm-fall: Practical and robust fall detection via mmwave signals," *IEEE Transactions on Mobile Computing*, 2025.

[113] J. Li, H. Xu, J. Zeng, W. Ai, S. Li, X. Li, and X. Li, "Radar-based human activity recognition using dual-stream spatial and temporal feature fusion network," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 60, no. 2, pp. 1835–1847, 2023.

[114] B. Cao, Q. Ping, B. Liu, Y. Nian, and M. He, "Real-time fall detection using wideband radar and a lightweight deep learning network," *IEEE Sensors Journal*, 2024.

[115] B. Wang, H. Zhang, and Y.-X. Guo, "Radar-based soft fall detection using pattern contour vector," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2519–2527, 2022.

[116] M. Gu, Z. Chen, K. Chen, and H. Pan, "Rmpct-net: a multi-channel parallel cnn and transformer network model applied to har using fmcw radar," *Signal, Image and Video Processing*, vol. 18, no. 3, pp. 2219–2229, 2024.

[117] X. Yuan, J. Li, Q. Chen, and G. Zou, "Mfecnet: Multi-feature fusion extra convolutional network based on fmcw radar for human activity recognition," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[118] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kernec, "Radar signatures of human activities," 2019.

[119] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, T. Macks, and F. Fioranelli, "Cross-frequency training with adversarial learning for radar micro-doppler signature classification (rising researcher)," in *Radar Sensor Technology XXIV*, vol. 11408, pp. 58–68, SPIE, 2020.

[120] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[121] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[122] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[123] Y. Qiu, X. Li, Z. Deng, X. Huang, P. Pan, and X. Ma, "A low-cost dual-path feature fusion network for omnidirectional human motion recognition using monostatic radar," *IEEE Transactions on Aerospace and Electronic Systems*, 2024.

[124] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PmLR, 2020.

[125] 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 *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, pp. 51–56, 2019.

[126] E. Gambi, G. Ciattaglia, A. De Santis, and L. Senigagliesi, "Millimeter wave radar data of people walking," *Data in brief*, vol. 31, p. 105996, 2020.

[127] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings*

of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.
[128] M. S. Seyfioğlu and S. Z. Gürbüz, "Deep neural network initialization methods for micro-doppler classification with low training sample support," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2462–2466, 2017.
[129] F. Jin, R. Zhang, A. Sengupta, S. Cao, S. Hariri, N. K. Agarwal, and S. K. Agarwal, "Multiple patients behavior detection in real-time using mmwave radar and deep cnns," in *2019 IEEE Radar Conference (RadarConf)*, pp. 1–6, IEEE, 2019.
[130] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.
[131] N. Nguyen, M. Pham, V.-S. Doan, and V. Le, "Improving human activity classification based on micro-doppler signatures of fmcw radar with the effect of noise," *PloS one*, vol. 19, no. 8, p. e0308045, 2024.
[132] 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 Aerospace and Electronic Systems Magazine*, vol. 37, no. 3, pp. 6–22, 2021.
[133] W. Yin, L. Shi, and Y. Shi, "Continuous human action recognition by multiple object detection based fmcw radar," *IEEE Transactions on Aerospace and Electronic Systems*, 2024.
[134] K. N. Faisal, H. S. Mir, and R. R. Sharma, "Human activity recognition from fmcw radar signals utilizing cross-terms free wvd," *IEEE Internet of Things Journal*, vol. 11, no. 8, pp. 14383–14394, 2023.

Van Ngoc Dang received his engineering degree in Control Engineering Technology and Automation from the Thai Nguyen University of Information and Communication Technology in 2013 and a Master degree in Control Engineering and Automation from Hanoi University of Science and Technology (HUST) in 2016. Currently, he is a Ph.D. research student at the School of Electrical and Electronic Engineering (SEEE), Hanoi University of Science and Technology (HUST). His research interests include Signal Processing and Machine Learning.

Ngoc Chau Hoang received the B.E. degree in control engineering and automation from the Hanoi University of Science and Technology, Hanoi, Vietnam, where he is currently working toward the M.Sc. degree in control engineering and automation with the School of Electrical and Electronic Engineering. His research interests include signal processing, speech enhancement, and machine learning theory.

Minh Thuy Le received her Engineer degree (2006) and M.S. degree (2008) in Electrical Engineering from Hanoi University of Science and Technology (HUST), and her Ph.D. degree (2013) in Optics and Radio Frequency from Grenoble Institute of Technology, France. In 2016, she conducted academic research at the mmWave Laboratory at the Nagoya Institute of Technology. From 2022 to 2023, she was a Visiting Professor at the University of Technology Sydney. Since 2013, she has been an Associate Professor and the Radio Frequency Group Leader at the School of Electrical and Electronic Engineering, HUST, Vietnam. Her research interests include antennas, indoor localization, energy harvesting, wireless power transfer, and smart sensors.

Kien Nguyen (SM'16) received a B.E. degree in electronics and telecommunication from the Hanoi University of Science and Technology (HUST), Vietnam, in 2004, and a Ph.D. degree in informatics from the Graduate University for Advanced Studies, Japan in 2012. Dr. Nguyen was a researcher at the National Institute of Information and Communication Technology (NICT), Japan during 2014-2018. Since 2018 he has been with Chiba University, where he is currently an associate professor. His research covers a wide range of topics in networking and distributed systems, including the Internet, the Internet of Things technologies, and distributed ledger technologies. His research achievements have been disseminated in three patents, several IETF Internet drafts, and more than 160 publications in peer-reviewed journals and conferences. He is a senior member of IEEE and a member of IEICE. He also participates in IETF activities.

Quoc Cuong Nguyen received his engineer (1996), M.S. (1998) degrees in Electrical Engineering from Hanoi University of Science and Technology (HUST), Vietnam, and Ph.D. in Signal-Image-Speech-Telecoms from INP Grenoble, France, in 2002. He is the Head of the Sensor Laboratory at the School of Electrical and Electronic Engineering (SEEE), Hanoi University of Science and Technology (HUST). His research interests include Signal Processing, Speech Recognition, Beamforming, Tinymachine Learning and Smart sensors.