Received 17 May 2024; accepted 5 June 2024. Date of publication 7 June 2024; date of current version 1 July 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3411529

WiFi-Based Human Sensing With Deep Learning: Recent Advances, Challenges, and Opportunities

IFTIKHAR AHMAD (Student Member, IEEE), ARIF ULLAH (Member, IEEE), AND WOOYEOL CHOI (Member, IEEE)

Department of Computer Engineering, College of IT Convergence, Chosun University, Gwangju 61452, Republic of Korea

CORRESPONDING AUTHOR: W. CHOI (e-mail: wyc@chosun.ac.kr)

This work was supported by the research fund from Chosun University in 2024.

ABSTRACT The rapid advancements in wireless technologies have led to numerous research studies that provide evidence of the successful utilization of wireless signals, particularly, WiFi signals for human activity sensing in the indoor environment. As a promising technology, WiFi-based human sensing can be utilized for a variety of applications such as smart healthcare, smart homes, security, industry, office indoor environments etc., due to the availability of rich infrastructure. Furthermore, compared to other radio frequency (RF) based systems such as radio detection and ranging (RADAR) and radio frequency identification (RFID), WiFi is non-invasive, has low-cost, and provides ubiquitous coverage in the indoor setup. However, due to the limited accuracy and high complexity of the model-based approaches for human sensing, the systems empowered by the deep learning (DL) techniques have achieved remarkable performance gains and showed more robustness in dealing with complicated human sensing tasks. The article explores the physical layer parameters used in WiFi sensing such as received signal strength indicator (RSSI) and channel state information (CSI), the estimated parameters such as angle-of-arrival (AoA) and Doppler shift (DS) along with frequency modulated continuous wave (FMCW) RADAR technology. Moreover, the preliminary signal processing stages that are applied to the received WiFi signals in the DL assisted systems are discussed. This article provides a comprehensive literature survey on the recent advances in DL-empowered WiFi sensing focusing on human activity recognition and movement tracking followed by fall detection, single task-multi task classification, crowd monitoring and sensing, indoor localization, gaits recognition, and pose estimation. Furthermore, the paper highlight the challenges in the existing literature and discusses the possible future research directions in WiFi-based human sensing assisted by DL techniques.

INDEX TERMS Deep learning, device-based sensing, device-free sensing, human activity recognition, human pose estimation, indoor localization, machine learning, RF-sensing.

I. INTRODUCTION

THE WIDE range of emerging wireless technologies has triggered significant advancement in the fields of wireless communication, networking, and sensing. Within this dynamic landscape, emerging radio frequency (RF) sensing stands as a notable and innovative stride that has the potential to redefine our perception and interaction with the world around us. By utilizing the basic principles of RF signals, RF-based human sensing systems are capable of identifying humans, detecting falls, monitoring humans, and more especially in the development of security, healthcare, and smart home systems [1], [2]. RF systems use radio waves to transmit and receive signals, and can be used for sensing by capturing significant information about the presence and movements of people within a defined space. This data is obtained by analyzing how the radio waves change as they interact with the humans in the surrounding. Unlike traditional methods that rely on cameras or physical sensors, RF-based sensing remains unaffected by visual obstructions, illumination, or external disruptions [3]. This characteristic offers benefit in scenarios where ensuring privacy, safety, and avoiding the necessity of a direct line of sight (LoS) is of utmost significance [4]. Human sensing technology can be either device-based and device-free. Device-based approach utilize dedicated hardware or wearable devices such as smartphones and motion capture systems to track movements while device-free sensing relies on existing infrastructure or ambient signals for human monitoring. One such technique utilizes RF signals for human activities recognition to enable non-intrusive monitoring in both LoS and non-line-of-sight (NLoS) scenarios. However, these sensing methods present unique advantages and challenges, which drive ongoing research for integration into healthcare, security, smart environments, and human-computer interaction.

A. DEVICE-BASED AND DEVICE-FREE SENSING

Within the realm of human-computer interaction and environmental monitoring, two primary paradigms emerge such as device-based human sensing and device-free sensing. The former involves the deliberate deployment of specialized sensors or devices designed to monitor and interpret human actions and behaviors such as wearable fitness trackers [5], motion sensors [6], pyroelectric infrared (PIR) sensors [7], stretch sensors [8], acoustic-based sensors [9], [10], smartwatches [11], [12], cameras [13], and medical monitoring sensors [14] that track different parameters like heart rate, steps taken, sleep patterns, and even skin temperature. On the other hand, Device-free human sensing uses RF signals to enables the observation and analysis of human behaviors, movements, and presence without relying on specialized sensors or wearable devices. This methodology takes advantage of existing infrastructure such as ambient signals, WiFi routers, and radio waves to detect alterations caused by human activities and motions within the vicinity [3]. To sum up, device-based sensing is remarkable for its accuracy in measuring specific data points but it can be invasive and expensive. On the other hand, device-free sensing though is not as detailed in data collection but it provides solutions that are respectful to privacy, is less intrusive, and could be scale up more easily by using the infrastructure already in place [15]. Choosing between these approaches depends on the particular situation and finding the right balance between accurate information and practical considerations. Device-free sensing is advantageous in a situation where the installation of physical sensors is challenging and to avoid expensive system expansion for effective monitoring.

B. RF-SENSING TECHNOLOGIES

RF-based human sensing constitutes a dynamic field that employs RF signals to comprehend and assess human presence and behaviors. Development in RF hardware, signal processing methods, and machine learning (ML) algorithms have collectively contributed the advancement to enhance accuracy, reliability, and real-time capabilities along with seamless integration of RF-based human sensing technologies. The deep learning (DL) has experienced notable progress in recent years driven by the accessibility of powerful computing resources such as GPUs, breakthroughs in training algorithms as well as growing interest across diverse domains [16]. The integration of RF-based sensing with DL has brought a transformative impact in contactless sensing and provide solutions to numerous challenges [17], [18]. A sophisticated artificial neural networks (ANN) characterized by multiple layers, commonly referred to as deep neural networks (DNN), is employed to extract vital information from extensive RF signals. Generally, the DNN comprise multiple hidden layers and follow iterative weight adjustments during data processing, thereby improve their learning ability and identify complex patterns to enhance their proficiency in extracting valuable information from RF data. DL models are adaptable to dynamic environments and improving the accuracy and robustness of real-time sensing applications by continuously updating their knowledge through learning from the RF data [19]. Additionally, DL enables the fusion of RF sensing with other sensing modalities such as cameras to improve the accuracy and insights in complex scenarios [20]. Furthermore, RF signal has inherent non-linear characteristics and DL models perform exceptionally in capturing these non-linear relationships between input features and target variables [21]. This nonlinearity is particularly crucial in RF sensing where basic linear models may find it challenging to comprehend the underlying patterns effectively.

Within the realm of RF sensing, sensing is done using radio detection and ranging (RADAR), radio frequency identification (RFID), and WiFi-based systems [22], [23]. RADAR utilizes radio waves to detect human movements and gestures and is applicable in domains like security and healthcare while RFID employs RF signals to recognize and oversee individuals via tags or cards and its application spans from access control to healthcare oversight. Meanwhile, WiFi-based sensing utilizes WiFi signals to detect and monitor human presence and collectively underscores the potential to enable a thorough comprehension of human behaviors and interactions. Expanding the discussion on RF-based human sensing methods, we provided brief overviews of RADAR, RFID, and WiFi-based sensing in the next sections.

1) RADAR-BASED SENSING

A typical RADAR system involves a transmitter and receiver, where an electromagnetic signal is transmitted by the transmitter and collected by the receiver with a processing capabilities to perform further data analysis. The RADAR systems can be classified into active and passive system. Both differs from each other in a sense, that active RADAR uses the bounce signal off the objects while passive RADAR analyzes disruptions in existing signals for detection. When objects intercept the transmitted radio signal results in scattering which propagates the information regarding the range, velocity, trajectory, and identity. The well known Doppler effect is utilized to measure the velocity by assessing frequency shifts in the RADAR signal. Object identification relies on micro-Doppler signatures, tiny modulations induced by slight movements in an object's oscillating or rotating parts. In human monitoring, these signatures reflect specific patterns in torso, limb, and head movements during motion unique to individuals and activities and serve automated classification and monitoring objectives effectively [24], [25].

Considerable research has been carried out in the area of RADAR-based human activity recognition (HAR) by utilizing DL approaches [26], [27], [28]. The authors in [29] present a novel method for hand gesture recognition using feature cubes encoded by three dimensional fast Fourier transform (3D-FFT) using convolutional neural network (CNN) architecture with specialized spatiotemporal convolution blocks. The approach achieves recognition rates of 99.53% and 97.22% on the dataset validation and testing stage, respectively, and surpasses the existing approaches. FallWatch is an learning model that detect real-time falls despite visual obstructions [30]. It employs trained CNN, attention mechanism, recurrent neural network (RNN), and long short-term memory (LSTM) fall detection from signal collected through antenna array. FallWatch excels in detecting falls across multiple individuals and settings, surpassing the alternative approaches, and offering a unique solution for elderly fall monitoring. Frequency-modulated continuous wave (FMCW) RADAR provides exceptional benefits with superior range and velocity resolution and is well-suited for tasks like precise distance measurements, gesture recognition, and effective object tracking [31], [32], [33], [34]. The authors in [35] employ FMCW RADAR for falls detection achieving 95.5% accuracy in identifying six fall-like movements using the dynamic range-Doppler trajectory (DRDT) and a subspace K-nearest neighbor (KNN) classifier. Similarly, a new preprocessing method is introduced in [36] to tackle static interference and estimate range, velocity, and angle features for gesture recognition using CNN.

2) RFID-BASED SENSING

Different from the traditional RFID applications concentrated on object identification, RFID-based human sensing employs RFID technology to monitor human presence and movements within the vicinity. This method prioritizes human interactions with the environment through strategically positioned RFID tags or sensors. The system recognize the entry of individuals with RFID tags, enabling presence detection, location tracking, motion analysis, and interactions with tagged objects. Its applicability spans smart homes, healthcare, and retail, fostering context-aware systems, occupancy monitoring, and personalized experiences derived from human behavior. The RFID technology is utilised for automatic identification in the context of the Internet-of-Things (IoT) technology [37], [38], [39]. Recent advancements in wireless sensing for tracking human activities have encountered challenges such as direct contact or specialized hardware requirements of RFID-based systems. The authors in [40] introduce TACT, a contact-free approach utilizing common RFID systems to detect and classify human activities and achieves 93.5% precision in diverse scenarios by segmenting and classifying phase values. The authors in [41] consider an ambient framework for HAR using multivariate Gaussian modeling and employs maximum likelihood estimation for feature learning. Comprehensive experiments shows accurate predictions, making it suitable for single as well as multi-dwelling settings.

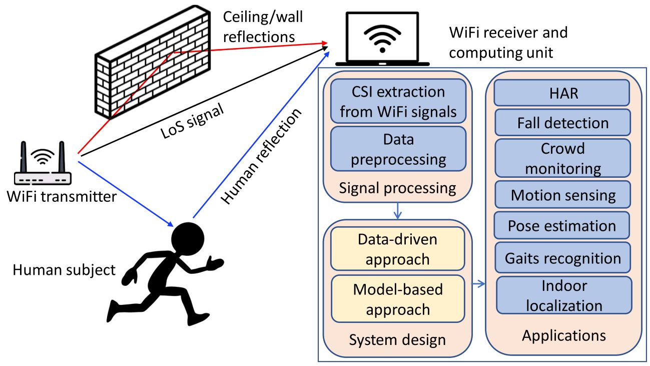

FIGURE 1. WiFi-based human sensing [3].

DL for RFID-based human sensing is explored in the literature such as the authors in [42] propose TagFree which is a pioneer device-free RFID-based HAR system considering the impact of multipath signal reception, correlations between signal power, angles, and activities, and capturing angle data while analyzing it through DL. TagFree is demonstrated by an Impinj reader prototype and show superior performance in multipath-rich environments. A passive RFID sensor tag system employing RFMicron's Magnus S chip is presented in [43]. It measures RSSI and pressure changes in indoor footwear during activities. Extracted features are used for ML-based activity classification and achieve high accuracy for different subjects. The authors in [44] introduce DeepTag which is an advanced DL-based RFID framework which extracts key features from phase data and uses a hybrid architecture for signal analysis, adapting well to various scenarios and demonstrating superiority in multipath-rich settings. The back scatter RFID tag can be read by WiFi device for sensing. However, tag needs to be deployed in the close proximity of WiFi device to work robustly in NLoS scenarios which limit their operating range.

3) WIFI-BASED SENSING

WiFi sensing is an emerging technology that uses WiFi signal to perceive and interpret the surrounding environment and offer numerous applications such as indoor localization, HAR, environmental monitoring, pose estimation, and even gesture identification. Fig. 1 illustrates the general framework for WiFi based human sensing and its applications.. Nevertheless, striving to interpret WiFi signals and derive substantive insights from them poses a major challenge. Compared to RADAR and RFID-based sensing, WiFi-based human sensing offers distinct advantages such as utilizing existing infrastructure and reducing deployment complexity and costs. Moreover, it provides higher resolution data due to higher frequency signals, enabling finer movement tracking. Unlike RADAR technology, which is unable to see through walls, WiFi-based sensing provides extensive coverage with no blind spots. Additionally, WiFi affordability and widespread availability enable WiFi sensing to detect presence in the WiFi range. Similarly, as RFID deployments are constrained by infrastructure limitations, WiFi sensing utilizes the existing WiFi infrastructure with minor firmware updates. Another advantage of WiFi over RFID is its superior accuracy, attributed to its wider bandwidth and more sophisticated signal processing capabilities, facilitating precise sensing [45]. The widespread WiFi adoption enhances scalability for smart environments, making it a versatile approach for applications like smart homes and healthcare. WiFi-based sensing capitalizes disruptions within the wireless channels caused by human activities and movement within the vicinity changes the signal propagation, manifesting as variations in channel state information (CSI) [46]. This dynamic CSI holds valuable insights into human motion and interactions. Leveraging the WiFi signals from WiFi routers, Intel 5300 network interface cards (NIC) [47], and Atheros [48] stands out for its cost-effectiveness and easy deployment. These devices integrated into WiFi infrastructure utilizes orthogonal frequency division multiplexing (OFDM) by dividing the signals into subcarriers with independent data transmissions [49]. Human presence disrupts subcarriers and alters signal attributes captured by the CSI.

WiFi devices operating in 2.4GHz or 5GHz bands provides balanced coverage and data rates for indoor and extended-range user. Extracting patterns from OFDM subcarriers, WiFi sensing identifies and categorizes human activities for context-aware applications. Different from conventional methods relying on received signal strength indicator (RSSI) which struggles with noise and subtle movements, the CSI approach exploits signal amplitude and phase details to detect minute motions at sub-cm levels across various frequency channels. This sophisticated technique encompasses data collection, signal processing, and classification stages reminiscent of detection processes. The WiFi based sensing methods includes model-based and learning based approaches. Both approaches differs from each other in a sense that the former methods rely on physical model based signal propagation such as Fresnel Zone while the later uses trained model to approximate function and classify the input WiFi signal. The model-based approaches withstand environmental changes, however, they often fail to achieve significant performance improvements in case of the complicated human activities in coarse-grained applications [46]. Furthermore, it is extremely challenging to determine accurate mathematical to approximate correlations between complex human actions and variations in WiFi signals. Moreover, the model-based approaches suffers performance degradation in NLoS scenarios. In this regard, WiseFi system is proposed in [50], where the authors illustrate this limitation and achieve a median recognition accuracy of just 74.3% in case of the signal transmission through single wall. On the other hand, the learning approaches perform well as it enables processing of high dimensional data. Although a substantial amount of work has been conducted in this field, there is still significant room for further enhancement. This article deeply explores the existing works in the domain of WiFi-based human sensing employing sophisticated DL approaches.

C. DEEP LEARNING FORWIFI-BASED SENSING

DL takes on a crucial role, granting significant benefits to WiFi sensing [18]. DL-based WiFi sensing stands out from other RF technologies such as RADAR and RFID due to its use of the extensive and continuous signal data provided by existing WiFi networks. This rich data stream enables more precise detection of subtle human behaviors and enhances the granularity of movement tracking. In comparison, other RF methods are typically more specialized, and tailored for specific applications that may not demand such comprehensive data collection. WiFi-based sensing with DL enables accurate classification of specific body movements, relying on unique WiFi signal changes linked to human presence and activities [57]. The authors in [58] designed WiFi CSI-based framework with CP factorization, graph learning based optimized features, and used RNN for daily activity detection that work well without requiring human intervention. Different ML algorithms are applied on CSI datasets for motion detection in [59] and the results show that RNN provide superior performance in case of larger datasets. The existing literature extensively employed DL techniques for the HAR using WiFi signals for a diverse set of tasks such as fall detection, classifying different activities, monitoring human behavior, and identifying walking patterns (gaits), etc., [60], [61], [62], [63], [64], [65]. SenseFi [57], is an open-source benchmark and DL library for WiFi-based human sensing evaluated for different datasets. Furthermore, the authors in [66] presents the detailed investigation of DL techniques on WiFi-based HAR using CSI.

D. LITERATURE SELECTION FOR THE SURVEY

For the comprehensive survey, an organized approach was used to search for related research publications on WiFi sensing. A range of reputable databases were meticulously accessed, including Google Scholar, IEEE Xplore, ACM Digital Library, Science Direct, Springer, and Archive. A range of keywords including “RF human sensing,” “WiFi sensing”, “DL”, “CSI”, “fall detection”, “pose estimation”, “localization”, and “HAR” were employed to identify relevant literature. The primary focus is to review the recent research works on WiFi sensing with a specific temporal constraint ensuring the inclusion of papers published in the last five years to capture the most current developments in the field. The source diversity was a key aspect, including survey papers, high-impact journal articles from esteemed journals like the IEEE IoT Journal, IEEE Transactions on Mobile Computing, IEEE Transactions on Aerospace and Electronic Systems, and conference papers from prestigious

TABLE 1. How does this survey vary from other surveys on RF-sensing. LC, MC, and HC denotes low, medium, and high coverage, respectively, while C and LR denotes the DL coverage and the number of literature surveyed on the topic.

Ref.	Year	Activity						DL Reviewed		Scope and Brief Description
		Fall Detection	Crowd Monitoring	Motion Sensing	Indoor Localization	Pose Estimation	Gaits Recognition	C	LR	
[51]	2017	LC	-	LC	-	-	-	-	-	This survey presents indoor HAR using WiFi CSI signal and performance evaluation using LSTM models.
[22]	2019	LC	-	HC	-	-	LC	-	-	The study introduces RF-based HAR (RADAR, Wi-Fi, and RFID), and its contributions to healthcare-assisted living systems.
[52]	2019	LC	LC	LC	LC	-	-	LC	<10	The article presents device free HAR across the wall using WiFi CSI and addresses challenges and future research directions.
[53]	2020	LC	LC	MC	LC	LC	LC	LC	<15	The paper explore the rising interest in device-free and CSI based human gesture recognition with focus on model and learning methods.
[54]	2020	MC	LC	HC	HC	-	LC	LC	<10	The provides a review on wireless sensing system followed by applications, challenges and potential future research directions.
[3]	2020	LC	-	LC	HC	-	LC	LC	<10	The survey presents WiFi-based human sensing, human detection and recognition using CSI and its applicability in a diverse applications.
[17]	2021	LC	LC	LC	MC	LC	LC	HC	83	This survey provides a detailed overview of DL in RF-sensing, publicly available datasets, discusses current challenges and future directions.
[55]	2022	LC	LC	MC	LC	HC	MC	MC	<30	This survey presents an innovative taxonomy of wireless device-free human sensing (WDHS) systems into eleven groups based on sensing tasks and motion details, challenges and future research directions.
[56]	2023	LC	-	MC	MC	LC	LC	MC	<30	The paper presents a comprehensive review on utilizing CSI for healthcare applications, highlight the current limitations and trends.
This survey	2023	HC	HC	HC	HC	HC	HC	HC	>100	This work present a comprehensive review on DL-based WiFi sensing, its applications, RF sensing parameters, signal processing, and challenges with future research directions.

events such as CVF, ICCV, ICML, ICASSP, and CVPR. We closely examined the literature to make sure it matches our research goals and keywords. By continuously Web searching over the databases, we ensured to identify the most up-to-date literature which led to a compilation of this work.

E. CONTRIBUTIONS AND PAPER ORGANIZATION

Numerous recent surveys have addressed RF-based human sensing, encompassing RADAR-based sensing [67], [68], RFID-based sensing [69], [70], and WiFi-based sensing [52], [53], [56], [71], [72]. However, these surveys provide extensive insights into diverse techniques and applications. This paper provides a literature review on WiFi-based HAR, emphasizing the integration of DL approaches with more comprehension and detailed analysis. Furthermore, WiFi-based HAR using DL techniques is deeply analysis with a specific emphasis on full-body-involved activities. It addresses a pivotal aspect of human behavior analysis, categorizing existing research, methodologies, and applications to enhance the understanding of leveraging WiFi signals for complex human activity detection. Moreover, a systematic presentation on the recent advancements, challenges, and opportunities is highlighted to shed light on employing DL techniques to improve accuracy, robustness, and real-world applicability. This work provides a guideline for the research community to advance research, innovation, development, and recent insights into WiFi-based HAR. To the best of our knowledge, this is the first survey that presents DL-aided WiFi-based HAR, particularly concentrating on whole-body activities. Table 1 provides a comparison of this work with other related surveys published on the topics.

The main contributions of this survey are as follows.

- This survey offers an extensive review of WiFi-based human behavior recognition through DL, highlighting contributions from a wide range of recent research works.
- This survey illustrates how DL effectively tracks human movements via WiFi signals, showcasing diverse methodologies and results.
- This survey presents studies that merge WiFi signals and DL for human pose estimation (HPE), illuminating the realm of contactless pose inference.
- This survey paper explores the existing challenges and possible future research directions in the realm of WiFi-based sensing.

As depicted graphically in Fig. 2, the rest of the paper is organized as follows. Section II explores the RF-based human sensing including the technical background, key concepts of RF sensing including WiFi, and the employment of DL for WiFi sensing followed by signal processing

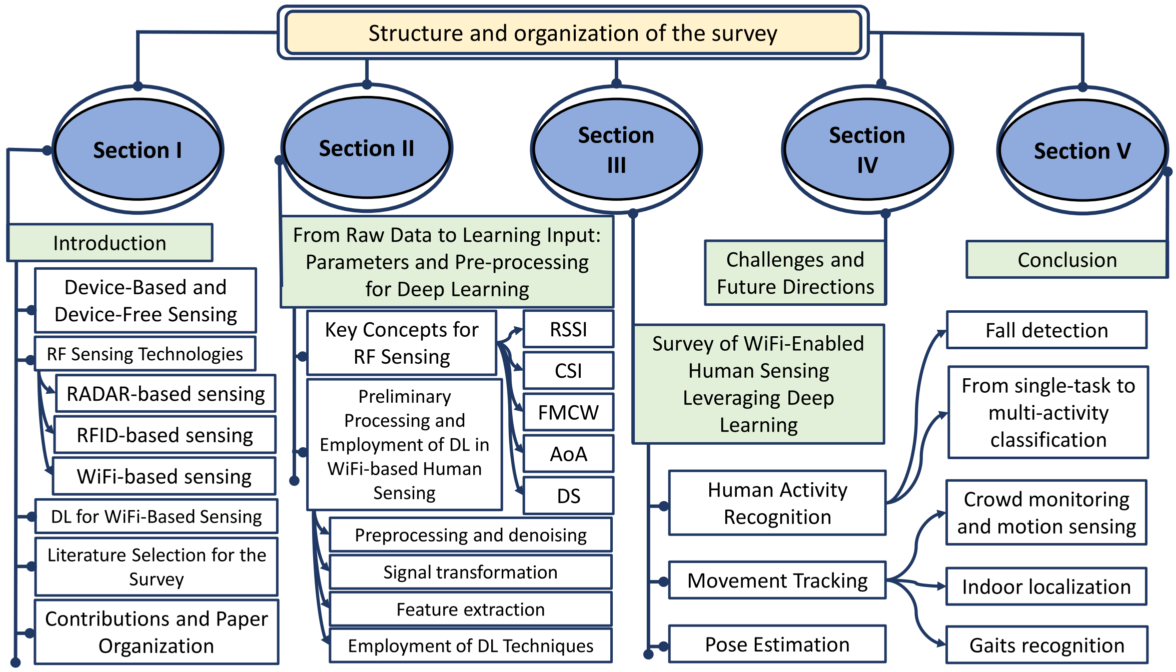

FIGURE 2. Paper structure and organization.

preliminaries. Section III provides a detailed overview of existing works on DL for different WiFi based sensing tasks. Section IV presents the challenges and the possible future directions related to WiFi sensing. Finally, the paper is summarized and concluded in Section V.

II. FROM RAW DATA TO LEARNING INPUT: PARAMETERS AND PREPROCESSING FOR DEEP LEARNING

For human activity detection, RF-based human sensing devices utilize technical aspects of received RF signals. The commonly used metric used for measuring signal properties in RF sensing includes RSSI, CSI, FMCW, angle of arrival (AoA), and Doppler shift (DS). This received raw signal is preprocessed, transformed, and significant features are extracted to decipher human actions accurately. These steps enable the subsequent analysis and recognition of diverse behaviors and play essential role in the initial stage of translating RF signals into valuable insights about human presence and motion. Generally, RSSI and CSI is utilized in commodity devices like WiFi while FMCW is used by RADAR and AoA and DS are estimated parameters.

A. KEY CONCEPTS FOR RF SENSING

In this section, we briefly discusses the physical layer parameters such as RSSI and CSI followed by FMCW RADAR technology, and other estimated parameters such as AoA and DS used in RF-based human sensing.

1) RECEIVED SIGNAL STRENGTH INDICATOR (RSSI)

The RSSI serves as a fundamental metric akin to a measuring instrument in RF sensing. It quantifies the amplitude of wireless signals propagating across varying distances. It finds extensive application in tasks such as localization of individuals [73]. Within the domain of human sensing, the proximity of an individual introduces signal attenuation which leads to dynamic fluctuations in the RSSI values in the sensing area. While readily accessible across numerous WiFi devices, RSSI's granularity remains limited, providing a singular numerical representation of signal loss [74]. This constrained information content hampers its capacity to discern a comprehensive spectrum of human activities. Nonetheless, the RSSI remains a valuable tool in the realm of RF human sensing, enabling insights into human movement through RF signal analysis [75]. Moreover, the RSSI values can exhibit variability even in the absence of movement which poses challenges to their reliability in real-world scenarios especially in fine-grained HAR.

2) CHANNEL STATE INFORMATION

The information regarding channel condition is a pivotal factor in RF human sensing, providing a valuable avenue to discern intricate human behaviors and motions by analyzing shifts within the wireless channel caused by these actions. Compared to metrics like RSSI, CSI offers a more detailed view of the wireless signal. CSI is composed of complex values that include both amplitude and phase information across multiple OFDM subcarriers. These subcarriers capture distinct multi-path fading effects and collectively create a detailed picture of the wireless channel. The IEEE 802.11n standard provides CSI measurements for 52 and 128 subcarriers, each with bandwidths of 20MHz and 40MHz, respectively, while the emerging 802.11ac standard supports even wider bandwidths [52]. As individuals move within WiFi or RF device coverage, their motion change to the wireless channel condition which leads to variations in the signal amplitude and phase across distinct subcarriers. Devices equipped with multiple-input-multiple-output (MIMO) transmission capabilities can capture these variations to construct the CSI matrix. The CSI matrix captures temporal and spatial changes due to human actions, enabling recognition of subtle gestures, actions, and even breathing patterns. DL techniques and pattern recognition algorithms learn the intricate patterns within CSI data and enable the identification and classification of specific human activities. Similar to RSSI, CSI measurements can be acquired using readily accessible WiFi devices such as Intel 5300 NIC [47], and Atheros [48] with customized drivers. The applicability of CSI-based human sensing spans security, healthcare, and smart homes, offering notable accuracy and versatility. Nevertheless, challenges exist such as advanced hardware, complex signal processing, environmental influences, and calibration requirements. In essence, CSI emerges as a robust asset in WiFi-based human sensing providing huge data that empowers precise and comprehensive HAR in diverse real-world scenarios through accurate analysis based on DL and advanced signal processing techniques.

3) FREQUENCY MODULATED CONTINUOUS WAVE

The FMCW is a vital technique in RF especially RADAR based human sensing which involves the emission of a continuous wave signal with a frequency sweep or chirp that gradually increases or decreases. This signal interacts with the environment and produces a frequency beat arises from the difference between the transmitted and received frequencies. This beat frequency carries crucial distance-related information about objects, facilitating accurate distance measurement [2]. FMCW excels in RF sensing because changes in the beat frequency due to human movements enable the detection and tracking of individuals' positions [30], [31], [76], [77]. FMCW's capacity to distinguish between various targets based on motion characteristics allows simultaneous identification of multiple individuals. Despite specialized hardware requirements, FMCW can generate and analyze the modulated signal at both the transmitter and receiver.

4) ANGLE OF ARRIVAL

The concept of AoA in RF-based provide a distinct approach to understanding human movements which determine the direction from which wireless signals reaches at the receiving antenna, offering insights into spatial orientation and movement patterns. Analyzing signal angles from different antennas make it possible to track activities such as walking, gestures, and posture changes [78]. Commonly, an antenna array captures signals from various directions where variations in the signal phase and intensity reveal arrival angles. TAoA data for human activity insights is derived through DL and signal processing. Advanced signal processing algorithms are required in such systems because the environmental factors such as signal reflection and interference effect the AoA of the received signal. Essentially, AoA enriches human sensing techniques and enables interpreting movement patterns in real-world scenarios using specialized hardware setups such as Texas Instruments FMCW RADAR sensors [17].

5) DOPPLER SHIFT

DS also known as Doppler effect refers to the change in frequency or wavelength of a wave due to the relative motion between the source and the observer. The DS has a significant role in RF sensing, offering a unique perspective to observe human activities and movements. Emerging from the changes in frequency caused by human motion in the vicinity, the DS helps identify and track scenario-dependent dynamics and imparts essential insights into movement speed and direction. Analogous to signals emitted by a transmitter and reflecting off the human body, any bodily motion triggers DS. Importantly, moving towards the receiver induces a positive frequency shift, while moving away causes a negative frequency change which facilitates the tracking the inference of motion patterns, covering activities like walking, running, and even vital signs [79], [80], [81]. The application of Doppler-based human sensing is especially crucial in domains requiring accurate real-time motion tracking, including security and healthcare applications. Despite its significance, effectively utilizing DS demands customized hardware setups and advanced signal processing to precisely extract and interpret frequency alterations [82].

B. PRELIMINARY PROCESSING AND EMPLOYMENT OFDEEP LEARNING IN WIFI-BASED HUMAN SENSING

Proceeding with attention directed toward the utilization of WiFi signals for human sensing, this section primarily focuses on CSI and RSSI. As a pivotal part of WiFi-based human sensing, signal processing techniques convert the raw signal data into valuable insights defining human activities. These methods encompass denoising for cleaner CSI data, signal transformation like Fourier and wavelet transforms to unveil temporal and frequency patterns, and feature extraction to distill these patterns for behavior recognition as shown in Fig. 3. Together, these strategies enable a precise understanding of human presence and motion in the WiFi coverage area.

1) PREPROCESSING AND DENOISING

In WiFi-based human sensing, preprocessing of CSI data is an essential phase that refines raw signal which involves a sequence of operations intending to improve the quality and reliability of collected CSI data. Initially the data calibration is performed to rectify systematic errors and ensure precise measurements. Subsequently, techniques like filtering and smoothing are applied to counteract unwanted interference, environmental fluctuations, and noise reduction [83], [84], [85]. The preprocessing phase also identifies

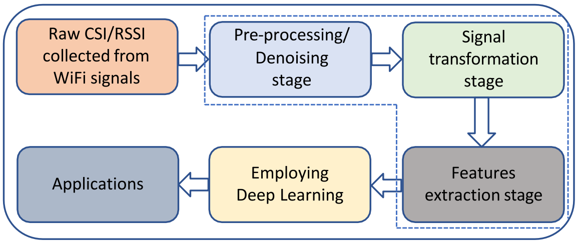

FIGURE 3. An overview of the signal processing sequence.

and eliminates outliers and anomalies in the data to further enhance the quality of the utilized dataset [86], [87]. Ultimately, preprocessing establishes a solid foundation for subsequent signal processing and DL techniques.

2) SIGNAL TRANSFORMATION

After preprocessing of WiFi signals, the collected CSI data requires signal transformation which is pivotal in revealing underlying patterns, temporal dynamics, and frequency characteristics within the data to enable more profound understanding of human activities and movements. One notable technique is the Fourier transform (FT) which translates the CSI data from the time domain to the frequency domain, unveiling the distinct frequency components embedded in the signal. This translation aids in identifying patterns and periodicity tied to specific human actions like walking and breathing. An additional powerful method is the wavelet transform which differs from the FT and allows both frequency and time localization [56], [88]. This capability is particularly useful for capturing rapid changes and transient events in the signal. Wavelet analysis furnishes a multi-resolution representation of the CSI data and facilitates the detection of both high- and low-frequency changes across varying time intervals to enable the recognition of a diverse array of human activities. Visual depictions like spectrograms or heatmaps assist in discerning frequency patterns correlated with human actions. However, in spite of the robust techniques, proper parameter adjustment and function selection are imperative to unveil pertinent information in the CSI data.

3) FEATURE EXTRACTION

After noise removal, certain systems directly derive features from the time-domain data, while others transform it into the frequency domain using techniques like short-time Fourier transform (STFT) or discrete wavelet transform (DWT) to create spectrograms. After preprocessing and signal transformation step, the features are extracted from the actual data in the next phase. The transformed data is translated into the representative features that encapsulate crucial insights about human activities and movements, forming the basis for subsequent analysis and classification. Feature extraction methods aim to capture pertinent aspects of the data while reducing noise and redundancy. Common techniques used for feature extraction include statistical features such as mean, standard deviation, skewness, and kurtosis which shows the characteristics of data distribution [89]. Temporal features analyze patterns within specific time intervals and uncover rhythms and periodicity. Frequency-domain features expose spectral components such as peak frequencies and spectral bandwidth, aiding in distinguishing activities with distinct frequency attributes. Time-frequency features such as mel-frequency cepstral coefficients (MFCCs) combine time and frequency information to capture nuanced variations [90]. Principal component analysis (PCA) and independent component analysis (ICA) are utilized for data analysis and dimensionality reduction [91]. They are commonly applied in ML to preprocess data before employing classification. The careful selection of relevant features is pivotal for ensuring the precision and effectiveness of subsequent analyses.

4) EMPLOYMENT OF DL MODELS FOR HUMAN SENSING

After applying the aforementioned signal processing steps on the collected dataset for a particular environment and sensing scenario, the DL model is trained to associate specific signal patterns with corresponding human activities such as walking, sitting, falling, pose estimation, localization etc. Once properly trained on large data, the model is tested on unseen data to evaluate the accuracy performance of the model. The trained model is then deployed to accurately predict and classify human activities by leveraging the learned parameters from WiFi signal data. Additionally, the DL models continuously learn and adapt to new patterns over time through techniques such as transfer learning to enhance the accuracy and robustness.

III. SURVEY OF WIFI-ENABLED HUMAN SENSING LEVERAGING DEEP LEARNING

In the recent years, DL has emerged as a powerful tool in various revolutionizing fields by replacing the traditional model based approaches with learnable model for data analysis and processing. Different DL architectures have been utilized in the literature for WiFi-based human sensing including multilayer Perceptron (MLP), CNN, simple RNN (SRNN), LSTM, BiLSTM, GRU, residual network (ResNet), autoencoder, transformer, and hybrid models. The hybrid model are the combination of multiple standalone DL models to enable efficient learning.

MLP: MLP comprises multiple interconnected layers of neurons with weighted connections. The MLP layers includes input layer, several hidden layers, and an output layer. Input data is fed in to the input layer and passed through the hidden layer employing nonlinear activation functions to process input data. Each hidden layers follow the data weighting and biasing before passing it to the output layer. This MLP architecture is trained via back-propagation and can be used for classification, regression, and feature learning.

CNN: CNN is feed forward neural network and process 2D grid data such images and matrix with correlated row and columns to learn spatial features in the input data by utilizing set of kernels for convolution [162]. CNNs

TABLE 2. Deep learning techniques for WiFi based sensing.

Category	Fall Detection	Single Task-Multi Task	Crowd Monitoring and Motion Sensing	Indoor Localization	Gait Recognition	Pose Estimation
MLP	-	[92], [93]	[94]	–	[93]	–
SRNN	[58]	[95], [58], [27]	[59], [96]	[97], [98]	[99], [100]	[101], [102], [103], [104]
LSTM	[30], [105], [106]	[107], [108], [44], [109], [95] [110], [60], [111], [30], [112] , [113], [114]	[115], [116], [117]	[98]	[99], [118]	[119], [120]
BiLSTM	–	[110], [60], [121], [114], [122]	–	–	–	–
GRU	[123]	[124], [123]	–	[125]	–	–
CNN	[105], [126], [30], [64]	[109], [127], [127], [62], [60], [18], [129], [31], [19], [130], [110], [131], [76], [121], [44], [113], [110], [111], [92], [114], [122], [30]	–	[132], [133], [134]	[118], [135], [136], [36]	[101], [119], [120], [102], [31], [137], [138], [139], [140], [142], [77]
ResNet	–	–	–	–	[135]	[104], [137], [142], [143], [144], [145], [146], [147], [146]
AE	–	[31]	[115]	[97], [148], [149]	–	[31]
Tranformer	[150], [151]	[150], [152], [151]	–	[153]	–	[154]
Hybrid	[159], [30], [63], [155], [156]	[92], [25], [95], [23], [32], [61], [20], [34], [121], [28], [44], [60], [113], [107], [111], [110], [114], [122],	[59], [115], [157], [158], [159], [94], [28]	[28], [21], [97], [98]	[99], [118], [135]	[101], [119], [120], [104], [31], [160], [161], [137]

uses supervised learning to extract features from the input data autonomously without human assistance, making them useful for applications like object identification and image classification.

ResNet: ResNet is a type of CNN having skip connections or shortcuts allowing to bypass certain convolutional layers. These connections empower ResNet to maintain high efficiency and effectiveness in training utilizing hundreds of layers. The core idea behind ResNet is to enable the training of the networks by allowing gradients to flow through the network without vanishing or exploding. ResNet utilizes identity mappings as skip connections and perform element-wise addition operations between these connections. ResNet is effective in tasks such as image classification, object detection, and segmentation.

SRNN: RNN models are exceptional at processing sequential data by utilizing their unique architecture to recall previous inputs, thereby influencing output decisions. RNN can do tasks that depend on the timing and sequence of data points because of their intrinsic ability to update their states continuously. Unlike other neural networks, which usually interpret input data independently, RNNs dynamically handle temporal dependencies [163]. The conventional simple RNN (SRNN) exhibit vanishing gradient problem which are dealt by other variants of RNN such as LSTM, BiLSTM, and GRU.

LSTM: LSTM model comprises multiple cells and gates. The cell remember the previous values and the three gates such as input gate, output gate and a forget gate regulate the data into and out of the cell. LSTM retain the information across long sequences and effectively deal with the tasks demanding comprehensive long-term contextual learning [128].

GRU: Gated Recurrent Units (GRUs) is another variant of RNNs which resemble LSTMs in operation but having a simpler architecture which comprises two gates instead of three, i.e., reset and update gate [164]. These gates empower GRUs to strike a balance between retaining past information and integrating new input, thereby simplifying the model while still effectively handling short-term memory challenges in sequences.

BiLSTM: BiLSTM is a modified version of LSTM which utilises two LSTM layer in a bidirectional way. In BiLSTM, the forward data flow is handled by one LSTM while the additional LSTM enable backward data propagation to model the two way sequential dependencies in the input data. The output of the two LSTM is then combined either by averaging, summing, or concatenating to generate the final output.

Autoencoder: Autoencoders is a specific type of DNN commonly used for unsupervised learning tasks involving dimensionality reduction [165]. Autoencoders comprise two main functions, i.e., encoder and decoder. The encoder compress the input data and the decoder reconstruct the original data from the compressed representation of the input data. Notably, the hidden layers of autoencoders typically have fewer nodes compared to the input layers, resulting in a bottleneck phenomenon that encourages the network to learn a concise representation of the input data. The autoencoder applies to data compression, enhancement, sophisticated data analysis, image processing, and noise reduction.

Transformer: The transformer is a powerful architecture based on the attention mechanism, famous for its ability to capture long-range dependencies in sequential data with remarkable efficiency and reduced training time [151]. A transformer comprises an encoder-decoder structure, where both components have multiple identical layers. In the encoder, each layer consists of two sub-layers: a multi-head self-attention mechanism and a fully connected feed-forward network, each followed by residual connection and layer normalization. The decoder includes an additional sub-layer for multi-head attention over the encoder output.

The brief discussion of these model is provided in the following paras while the application to different sensing scenario is listed in Table 2. In the next sections, we present the detailed survey of the aforementioned DL technique applied to WiFi based human sensing in the state-of-the-art literature.

A. HUMAN ACTIVITY RECOGNITION

WiFi-based HAR is an emerging and increasingly popular approach in recent years which non-intrusively monitor and analyze human behaviors across diverse environments. With continuous advancements in signal processing, neural network models, and deployment of wireless device, WiFi HAR shows great promise in applications such as healthcare, smart environments, security, etc. It provides robust healthcare monitoring through gait recognition and improves eldercare safety by detecting fall. The domain of indoor environments has advanced to incorporate motion sensing and population monitoring, along with precise indoor localization that is essential in huge complexes such as airports and malls. Pose estimation assists treatment, increases worker safety in factories or workplaces through monitoring, and enhances online shopping experiences with virtual try-on elements. These developments contribute to a deeper understanding of human interactions and ultimately enhance the overall quality of life. This section provides a discussion on the advancements and comparison of the state-of-the-art related to WiFi-based HAR including fall detection, multi-activity classification, movement tracking, indoor localization, gaits recognition, and HPE. Table 3 provides a comprehensive performance assessment of various learning models for activity recognition in terms of activity types, data collection sources and metrics, antenna systems, and key performance indicators such as precision, false alarm rate (FAR), accuracy, specificity, sensitivity, recall, and F1-score.

1) FALL DETECTION

Human fall is a major problem among elderly citizens worldwide which results in a considerable number of fatalities, particularly for people aged 60 and above. According to the World Health Organization (WHO), approximately 646,000 fall-related deaths occur globally and there are around 37.3 million falls that are treated medically each year [172]. It is projected that by 2030 the number of aged people (60 years or above) will comprise one-sixth of the global population. It became imperative to ensure timely rescue for elderly individuals living alone which presents a pressing public health challenge [173]. To address this issue, traditional methods have been utilized in the literature such as cameras, wearable sensors, and specialized RADAR hardware to detect falls. However, WiFi-based solution have gained popularity due to the widespread availability and non-intrusive nature. Utilizing advanced technology such ad MIMO-OFDM-based WiFi systems offers a promising approach to detect falls without human intervention and dependence on lighting condition requirements.

Extensive researchers has been carried out to improve fall detection in real-world environments using DL algorithms and WiFi based system. These efforts enhance the capability to accurately distinguish falls and ensure the safety of individuals, especially in challenging scenarios where traditional methods have limited performance. Authors in [155] proposed a DL method based on EfficientNet [156] for fall detection utilizing CSI data collected from WiFi signal and achieve over 96% accuracy. FallCNN is proposed in [126] which is a CNN based learning framework for fall detection utilizing CSI from WiFi signals. The proposed scheme is capable of achieving an average accuracy of 95% in seven different indoor environments. Similarly, the authors in [106] investigate fall detection in construction site using different variants of LSTM and reported 99% accuracy on the study case involving six workers and 360 activity sets. The authors in [166] proposed WiFall which utilizes WiFi CSI to detect human falls without requiring additional hardware, environmental setup, and wearable devices. Extensive simulations shows that WiFall can accurately detect falls for a single person in the coverage area. Using a single class support vector machine (SVM) classifier, the system achieves an accuracy of 90% with an average FAR of 15% across all testing scenarios while employing the random forest algorithm, WiFall achieves an average fall detection precision of 94% with FAR of 13%. RT-Fall is presented in [130] is a cost-effective indoor fall detection system that utilizes fine-grained CSI including phase and amplitude from commodity WiFi device and achieves real-time contactless fall detection. The authors demonstrate that the phase difference of CSI between two antennas is a reliable signal for activity recognition which enable accurate fall segmentation. Moreover, a sharp power profile decline pattern in the time-frequency domain for falls is identified which extracts new features to improve fall segmentation and detection. Experimental results demonstrate that RT-Fall outperforms the state-of-the-art approaches for four indoor scenarios with an average increase of 14% in sensitivity and 10% in specificity.

Researchers strive to develop a fall detection system with minimal environment dependency and do not requires retraining with new data while operating in the new environment. In this regard, the authors in [105] proposes an adversarial data augmentation (ADA) for WiFi-based domain-independent fall detection system. The CNN-ADA and LSTM-ADA models are utilized and compared on the FallDeFi dataset including five environments data. The dataset is divided into 10 domains and the semantic distance between the target and the source is measured using Wasserstein distance parameter (p). It is reported that the accuracy of CNN-ADA drop from 64.12% to 41.51% while LSTM-ADA exhibits an opposite trend with performance improvement from 40.20% to 66.03% with an increase in p. However, both models shows poor performance for real-world applications indicating further room for improvement in accuracy and reliability. Similarly, the authors in [174], [175] propose DeFall which utilizes speed and acceleration patterns for fall detection and is capable of working in different environments. The system estimates fall

TABLE 3. WiFi-based human activities recognition.

Ref.	Activity Type	Data Collection Source/Type	Data Collection Metrics	Antenna syst. /Fre.	Learning Models	KPI
[166]	Fall detection	WiFi/Passive	CSI	MIMO/5GHz	SVM, RF	Precision 94% FAR 13%
[130]	Stand then fall Walk then fall	WiFi/Passive	CSI	SIMO/5GHz	SVM	Sensitivity 89% Specificity 93%
[105]	Fall detection	WiFi/Passive	CSI	SIMO 2.4/5GHz	CNN-ADA LSTM-ADA	Accuracy (%) CNN-ADA 64 LSTM-ADA 66
[127]	Squatting, walk, boxing, run	WiFi/Passive	CSI	SIMO 2.4/5GHz	LSTM, CNN	Accuracy (%) Recognition: 95 Segmentation: 84
[92]	Trip, slip, balance lose, kneel fall, chair fall, walk then fall	WiFi/Passive	CSI	MISO 5.825GHz	DNN FCs +VAE/D	FAR 5.7% MAR 3.4%
[167]	Fall in trip, slip, consciousness loss	WiFi/Passive	AoA	SIMO 5GHz	ABLSTM	Accur. 84.31% F1 Score 84.56%
[111]	Sit, stand, lie down, walk, stand from floor	WiFi and UWB /Passive	CSI and CIR	UWB:SISO 4GHz WiFi:SIMO 5GHz	CNN, KNN, RF, DNN, and GNB	F1-Score UWB 95.53% WiFi 92.24%
[121]	Walk, run, stand, sit, squat, kick, greet, hug, shake hands, fight	WiFi/Passive	CSI	SIMO 5GHz	SVM	Accuracy (%) Strenuous 91 Normal : 84.43
[108]	Sit down, jump, run, pick up, walk, wave	WiFi/Passive	CSI	SIMO 5GHz	CNN, EM (RF, SVM, MLP)	Accuracy (%) Room Act: 98.1 Lab Env.: 99
[129]	13 human daily activities	WiFi/Passive	CSI	SIMO 5GHz	LDA, SVM	FS+SVM outperformed LDA by 0.1 ER
[107]	Walk, fall, run, sit pick, push, wave	WiFi/Passive	CSI	MIMO 5GHz	CNN+LSTM	Accuracy 96%
[110]	12 human daily activities	WiFi/Passive	CSI	SIMO 2.4GHz	CNN-ABiLSTM	Accuracy (%) Env-1: 98.54 Env-2: 91.96
[168]	Fall, walk, run, sit down, stand up, lie down	WiFi/Passive	CSI	SIMO 2.4/5GHz	ABLSTM	Accuracy (%) Room 96.7 Meeti. room 97.3

(Continued)

speed using WiFi signals in the offline stage and generates representative templates using dynamic time warping (DTW) algorithms and compare the real-time motion with templates to detect falls in online stage. DeFall achieve a detection rate of 96% and a false alarms less than 1.5% through a single pair of WiFi transceiver tested in various scenarios. The

TABLE 3. *(Continued.)* WiFi-based human activities recognition.

[169]	Board wipe, move suitcase, move, walk chair rota, sit, stand	WiFi/Passive	CSI	MISO 2.4/5GHz	CNN, FCLs	Accuracy around 0.75
[170]	Lay, stand, walk, fall, empty	WiFi/Passive	CSI	SIMO 5.32GHz	CNN+LSTM	Accuracy (%) Env-1: 73.4 Env-2: 80.3
[171]	Empty, sit,walk, run, stand,	WiFi/Passive	CSI	SIMO 5.32GHz	CNN+LSTM	Accuracy (%) 1-shot 86.8 5-shot 93.4
[124]	Sleep, walk, run, sit, wave, drink, throw,	WiFi/Passive	CSI	MIMO 2.4/5GHz	BI-LSTM, BI-AT-GRU	Accuracy (%) Env. 1: 97.4 Env. 2: 93.3
[150]	Jump, walk, bow, run, sit down, wave hand	WiFi/Passive	CSI	MIMO 2.4/5GHz	THAT Transformer	Accuracy (%) Env. 1: 98.4 Env.2: 99
[151]	Fall, walk, clean, run, circle, boxing	WiFi/Passive	CSI	MIMO 5GHz	CaiT Transformer	Accuracy (%) NTU-FI HAR: 98.2 UT-HAR 98.78
[152]	Squat, stand up, bend, boxing, turn around	WiFi/Passive	CSI	SIMO 2.4/5GHz	FCGTN Transformer	Accuracy (%) 1-shot: $>$ 95

FallWatch system presented in [30], detects falls in open-air and through the wall using autoencoder with CNN based encoder and LSTM models based decoder. The FallWatch detect falls effectively in a cross-person and cross-people scenario by achieving precision, recall, and F1-score of 0.923, 0.917, and 0.920, respectively. The authors in [92] by propose FallDar for fall detection which outperformed the state-of-the-art approaches by simultaneously tackling the environmental diversity, motion diversity, and user diversity challenges. FallDar utilizes the body's speed to handle environmental diversity and ensure resilience to changing conditions. FallDar uses DNN-based generative model for efficient detection of different types of falls. For user diversity, fall detection network is augmented with a user identification network to extract independent user features without requiring new user fall data. FallDar is implemented on commercial WiFi devices and tested for over six months in home and office environment trials, achieving FAR and missed alarm rate (MAR) of 3.4% and 5.7%, respectively. Similarly, the authors in [167] propose AFall which is a model-based robust fall detection system that utilizes WiFi CSI without requiring prior training for individuals. AFall employs the MUSIC algorithm to model the relationship between human falls and changes in AoA of the reflected WiFi signal from the human body. By using two receivers in orthogonal spatial layouts, AFall captures diverse AoA information and maintains stable performance even with slight environmental changes due to the independence of AoA with the surroundings and subjects. AFall is evaluated in five indoor environments and achieves an average accuracy of 84.31% and an average F1 score of 84.56%. Despite these advancements, the WiFi-based fall detection system is still immature and requires further improvement in terms of effectiveness and reliability.

2) FROM SINGLE-TASK TO MULTI-ACTIVITY CLASSIFICATION

Professionals are developing WiFi-based multi-activity classification to enable natural human behavior recognition, a promising technology with the potential to enhance applications across these diverse domains such as healthcare, industry, and smart homes. The authors in [76] utilize a CNN with a rectified linear unit (ReLU) activation function to estimate six different human activities. The experiment involved generating signals in compliance with the IEEE 802.11ad standards using a 60 GHz channel and utilize the DS information extracted from pilot subcarriers at the input of CNN. The CNN performed impressively by achieving a validation accuracy of 96%, demonstrating its proficiency in accurate recognition of human activities and the substantial reduction in the loss function value up to 0.1599 indicates the network's effectiveness in terms of error minimization during training. The authors in [111] present a comparison between ultra-wideband (UWB) and commodity WiFi setups for passive LoS based HAR and evaluated the estimation performance for five different human activities, i.e., sitting, standing, lying down, standing from the floor, and walking using a range of classifiers including CNN, DNN, Gaussian naive Bayes (GNB), KNN, and random forest. The channel impulse response (CIR) samples is used as input to the classifiers for UWB system while CSI data is utilized for commodity WiFi system. Simulation results show that CIR-aided UWB system gives an outstanding F1-score of 95.53% and the WiFi CSI amplitude-based input data achieves lower F1-score of 92.24% while the WiFi spectrogram-based input data yields an F1-score of 80.89%. Additionally, irrespective of whether the UWB or WiFi systems are employed, the random forest classifier consistently outperforms the other classifiers. However, UWB technology presents increased complexity and diminished resilience in NLoS scenarios compared to commodity WiFi systems. A one-class support vector machine (OSVM) is used in [121] which highlights the potential of WiFi-based CSI data for precise multi-activity recognition in complex scenarios. The system distinguishes between a range of human activities including walking, running, sitting, standing, and greeting as well as strenuous motions such as fighting and kicking. The system achieves notable accuracy rates of 90.89% for strenuous human motions and 84.43% for normal human body movements. Similarly, the authors in [122] present a CNN-attention bidirectional long short-term memory (CNN-ABLSTM) based WiFi CSI passive sensing technique, which outperforms the existing learning models in terms of accuracy for multiple actions and multiple individuals. In [112] seven distinct human activities are classified using InceptionTime and LSTM-based classification models applied to WiFI CSI data. Similarly, a hybrid DL network CNN-gated recurrent unit-attention network (CNN-GRU-AttNet) is tested on CSI-HAR and StanWiFi datasets. The proposed scheme outperforms the traditional DL models with an accuracy improvement of 4.62% compared to GRU, bidirectional GRU (BiGRU), CNN, LSTM and BiLSTM [114]. The WiFi CSI data is translated into images and processed by CNN based ImgFi in [131], showing 99.5% accuracy for human gestures recognition. WiAReS is an innovative device-free that accurately identify various activities [108] by utilizing CNN to extract spatiotemporal features from CSI measurements by capturing the spatial and temporal patterns of human activities. For enhanced recognition accuracy, WiAReS introduces a unique ensemble architecture that combines multiple models including a multi-layer MLP, a random forest, and an SVM. The CSI data is collected using standard WiFi devices equipped with Atheros-CSI-Tool on AR9590-based WiFi NICs for activities such as sitting down, jumping, wave, pick up, walk, and run. Extensive experiments were conducted in activity rooms and laboratories to demonstrate the exceptional performance of WiAReS with an overall accuracy of 98.1% and 99%, respectively. Similarly, the authors in [31] use predictive approach with trajectory-guided unsupervised learning (TGUL) for the recognition of multiple human actions and demonstrate that the proposed scheme achieves a mean average precision (mAP) of 91.7 in the fine-tuning setting for the intersection union threshold of θ=0.1.

The existing research exhibits domain dependence sensing, where a system performing well in one environment and exhibits performance degradation when tested in another environment. To overcome this, a robust system is needed to accurately identify human activities in diverse environments. Utilizing DL algorithms, the researchers aim to extract activity-specific features from the signal while avoiding environment-specific characteristics and ultimately design a versatile scheme independent of the sensing domain. In this regard, the authors in [129] investigate WiFi CSI based framework to recognize 13 indoor human activities using a single transmitter point (TP) and access point (AP). Online filtering is used for smoother CSI curves, preserving patterns, and a segmentation technique is utilized to isolates primitive action patterns from MIMO signal. SVM-based multi-classification with selected features enables activity recognition regardless of location, orientation, and speed. The results show that the SVM with feature selection performs better than linear discriminant analysis. A WiFi-based activity recognition system called WiSDAR is presented in [107] which incorporates spatial diversity awareness and achieves an accuracy rate of approximately 96% in detecting various human activities including walking, falling, sitting, running, picking, pushing, waving, and boxing. The proposed WiSDAR addresses the accuracy degradation challenges that arises when individuals pass through dead zones with ineffective signal coverage within the monitored environment. This is accomplished by integrating spatial diversity techniques. WiSDAR uses a DL model that combines CNNs and LSTM to achieve enhanced HAR performance. The authors in [95] use CNN and hybrid model of RNN and LSTM for HAR and achieve an average testing accuracy of 97%. Similarly, two DL models including attention-based bidirectional long short term memory (ABiLSTM) and CNN-ABiLSTM are utilized in [110] to recognize different human activities across diverse spatial environments. The experimental results demonstrate that by employing a transfer learning technique, these models deliver acceptable results when applied to new environments with varying configurations. The performance of WiSDAR for passive HAR is further improved in [168] through the utilization of ABLSTM. The ABLSTM utilizes a BiLSTM architecture to extract meaningful features from sequential CSI measurements in both forward and backward directions. To capture the varying importance of these learned features, an attention mechanism is incorporated that assigns different weights to each feature. The experimental results demonstrate that the ABLSTM achieves superior recognition performance compared to benchmark schemes. Moreover, its robustness and generalizability capable them to maintain high performance even when trained in one environment and tested in a different one. Another environment-independent (EI) approach is presented in [169] which extract activity features independent of the environment and subject. The proposed approach employs a feature extractor, an activity recognizer, and a domain discriminator, where the domain discriminator predicts the environment and encourages the feature extractor to generate environment/subject-independent features. Experimental results on six testbeds validate the effectiveness of the approach in transferable learning features. Similarly, CSI-based Parallel convolutional networks-based location-independent HAR system (CSI-PCNH) is proposed in [113] which uses 3DCNN with channel attention mechanism (CAM) and 2DCNN along with LSTM to recognize six different activities across various indoor locations and achieve an accuracy of 91.7%. Eight different location-independent human activities are recognized in [109] using CSI data and Raspberry Pi 4 with an LSTM DL model, achieving an overall accuracy of 97%. The authors in [170] proposed a new approach called activity-related feature extraction, enhancement, and matching network (AFEE-MatNet) which combines activity-related feature extraction and enhancement (AFEE) with a matching network (MatNet). AFEE focuses on improving CSI quality by eliminating non-activity-related features while preserving behavior-related information and reducing feature signal size for faster training. MatNet learn transferable features that can be applied across different environments and effectively handle domain shift challenges. Furthermore, the approach also incorporate a prediction checking and correction mechanism to rectify classification errors that arise from the expected human behavior transitions. Experimental results demonstrate that AFEE-MatNet outperform the existing HAR methods in terms of accuracy and training time. Similarly, the authors in [171] propose matching network with enhanced channel state information (MatNet-eCSI) which uses one-shot learning approach for HAR that is operable and can identify various activities in different environments. It combines the correlation feature matrix (CFM) and MatNet architecture comprising CNN with ReLU activation and Max pooling. The CFM is obtained by extracting activity-related information using a linear recursive operation and subtracting it from the received signal using an exponentially weighted moving average (EWMA) approach to reduces the dimensions of the original CSI matrix. MatNet utilizes the CFM and employs the CNN along with the Bidirectional LSTM model to produce the final output. Experimental results demonstrate the exceptional performance of MatNet-eCSI which surpasses the existing sensing methods with an average accuracy of 0.868 for one-shot learning. Additionally, it offers the advantage of reduced training complexity making it highly effective solution for HAR. Similarly, the authors in [127] propose DeepSeg which employs a CNN framework that transforms activity segmentation tasks into classification task, showing a significant performance improvement. The authors in [151] use vision transformers such as vanilla ViT, SimpleViT, DeepViT, SwinTransformer, and CaiT for CSI-based HAR on UT-HAR and NTU-Fi HAR datasets. Similarly, cross-domain action recognition using WiFi (CDFi) is introduced in [152] that utilizes transformers for activity recognition with minimal CSI samples and achieve enhanced performance compared to the state-of-the-art techniques in both cross-user and cross-scene scenarios. The authors in [124] integrate bi-directional and attention mechanisms into the gated loop unit (BI-AT-GRU) network to recognize diverse human actions. This approach achieves recognition accuracy of 97.4% and 93.3% across two distinct environments. Similarly, deep gated recurrent unit (DGRU) model is introduced for non-intrusive HAR in [123] which achieve 95% to 99% accuracy for various daily human activities. Similarly, a novel dual-stream convolution augmented human activity transformer (THAT) is used in [150] to effectively captures spatial and temporal features for various human activities and outperform existing state-of-the-art models in both effectiveness and efficiency.

B. MOVEMENT TRACKING

Movement tracking is the process of monitoring and recording the paths that objects or individuals follow over a period of time. This technology has diverse applications, such as sports analytic tracking, healthcare monitoring, and security surveillance. Movement tracking may includes crowd monitoring and motion sensing, localization, and gait recognition.

1) CROWD MONITORING AND MOTION SENSING

Crowd analysis is a prominent research area and is applicable in urban planning, crowd management, surveillance, healthcare, and prevention of accidents and casualties. Vision-based methods for crowd analysis suffer from blind spots, high deployment costs, computational complexity, poor lighting issues, and privacy concerns. In contrast, WiFi-based crowd monitoring offers low-cost, extensive coverage without blind spots, low computation, and preserves public privacy. This section provides a thorough literature review on WiFi-based crowd counting and motion sensing. A novel idea is presented in [176], which accurately calculates the time-up and go (TUG) time for a subject rising from the rest. The precise determination of TUG time depends on both the aligned movements of the subject and the continuous analysis of the reflected RF signals from the individual. TUG time is a crucial parameter widely used in the healthcare domain. It helps healthcare professionals to diagnose disease and find its severity in an individual. The authors in [177] present an in-depth review of AI-based crowd-counting methods in healthcare, biotechnology, and in IoT. By using RNNs with spatial-temporal matrices and grid partitioning, the paper [96] addresses the scalability of WiFi crowd counting while lowering error rates in campus tests. The study in [115] aims to enhance real-time monitoring using bi-convolutional LSTM, attention-based autoencoders, and semi-synthetic datasets for improved WiFi-based crowd analysis.

The work in [157] demonstrates the simultaneous tasks of crowd counting and localization estimation through ML and WiFi CSI without necessitating user-device interaction. In three different span of time, the experimental results achieved a counting accuracy of up to 94% and localization accuracy of 95% for k-fold cross-validation. This work is further explored in [158] and proposed Wi-CaL which addresses the feasibility of a simultaneous crowd estimation system capable of predicting crowd numbers and locations using WiFi IoT CSI technology and ML techniques. Instead of a conventional WiFi system, the author employed Espressif system platform 32 (ESP32) nodes and their CSI toolkit for medium-scale crowd counting and localization to establish a novel CSI platform. The experimental results show that the Wi-CaL achieved promising results. Using ML evaluation, the proposed method achieves 0.35 median absolute error (MAE) and 91.4% accuracy for 5 individuals in a small room and 0.41 MAE and 98.1% accuracy for 10 individuals in a medium-sized room. Compared to the percentage of non-zero elements metric (PEM), Wi-CaL outperformed with 0.41 MAE and 81.8% while PEM scored 0.62 MAE and 66.5% accuracy in single-person scenarios. The authors in [117] introduced CROSSCOUNT, a precise human counting system, that uses recurrent neural networks while addressing challenges like weak signals, data collection complexity, and the absence of CSI data. The system processes WiFi link intervals, surpassing received signal strength (RSS)-based methods. Utilizing LSTM to classify blockage patterns, CROSSCOUNT outperforms RF-based systems in counting accuracy. Across diverse test setups, it achieves 59% precise human count accuracy, improving to 100% within a two-person difference case. Similarly, the authors in [178] aim to count people within buildings using external WiFi transceivers, relying solely on WiFi RSSI measurements. It observes the inter-event times that are related to signal dip events and remain stable even behind walls. A method is introduced to derive people count from these inter-event times, treating wireless power measurements as a blend of renewal-type processes. By leveraging concepts of the renewal process from the literature, the probability mass function of inter-event times is calculated to predict crucial people count information. The results demonstrate the accurate estimation of people behind walls without extensive prior calibrations. In [94], the authors introduce WiCount, an innovative system, that utilizes DL techniques to address complex multi-human environmental sensing tasks using WiFi CSI (amplitude and phase). WiCount employs a DNN model based on back-propagation (BP) which comprises an input layer, two hidden layers, and a softmax layer. WiCount is capable of accurately counting up to five individuals. The system incorporates both offline training and online testing phases. Initial findings indicate that WiCount achieves an impressive average recognition accuracy of 82.3% for scenarios involving up to five people and outperforms traditional methods like SVM that operate on conventional features.

The authors in [159] introduce a device-free crowd-counting technique utilizing WiFi signals. By analyzing the phase and energy of subcarriers in real-time CSI, a DNN model is established to correlate these features with crowd density to enable accurate crowd counting. The method achieves a mean counting error of 0.11-0.14 with over 99% accuracy within medium crowds and 100% accuracy within small crowds. This level of precision satisfies the requirements of most crowd-aware applications, accommodating varying crowd sizes within WiFi-covered areas. The paper [179] introduces FreeCount, a device-free crowd-counting method using WiFi routers. A novel OpenWrt firmware is used to capture router CSI data and enable accurate estimation of occupants using two routers. Wavelet denoising, diverse feature calculation, and info-theory selection enhance counting while transfer kernel learning ensures temporal robustness. FreeCount achieves 96% accuracy with a promising practical implementation. The paper [116] employs WiFi sensing and LSTM models to predict event crowd counting. The study involves monitoring an event in Brussels using privacy-preserving WiFi sensors to gather crowd counts. These counts are transformed into time series data which is then used for forecasting. The research evaluates five LSTM models for crowd forecasting and compares their performance to a random walk model. The convolutional LSTM model stands out as the most effective and concludes the effectiveness of this system for accurate crowd monitoring.

IMep is proposed in [180], a system that operates without devices and utilizes WiFi signals to count the steps of multiple people. By utilizing CSI data, the authors developed the multiplayer amplitude decomposition algorithm (MADA) using block term decomposition (BTD). They also constructed the multiplayer stepping amplitude relation model (MSARM). In contrast to traditional single-step measurement methods, the study proposed a new moving energy method (MEM) technique to enhance step counting accuracy, leading to heightened precision in IMep's results. Experimental results demonstrated IMep's capacity to accurately count the steps of up to seven individuals simultaneously in real-world scenarios. Remarkably, IMep achieved step counting accuracy of 95.57%, 94.66%, and 89.94% in three different environments while also displaying adaptability to environmental changes. The authors in [181] propose WiStep, a WiFi-based step-counting method, leveraging multipath propagation. WiFi signals are modulated by the differential movement speeds of limbs and torso during walking and introduce distinct frequency components into received CSI. The approach employs time-frequency analysis to segment walking patterns, selecting sensitive subcarriers based on amplitude variance, and applying wavelet decomposition for faster processing. Step counting uses short-time energy of coefficients with combined results from chosen subcarriers for accuracy. WiStep accommodates in-place walking and exhibits resilience across diverse scenarios, achieving step counting accuracies of around 90% in experiments using commodity WiFi devices. The authors in [182] introduce Widar, a unique tracking system that utilizes WiFi signals for accurately estimating movement speed, direction, and locations of individuals at a precision of a few decimeters. Unlike conventional statistical learning methods, Widar establishes a geometric theoretical model to quantitatively

TABLE 4. Whole body movement tracking.

System/ Ref.	Tracking Activity	Experimental Setup		Learning Model	KPI
		Source/Type/ Measurement.	Antenna System /Freq		
Wi-CaL [157], [158]	Crowd counting and location estimation	ESP32 WiFi node /Passive/CSI	MIMO/2.4GHz	DNN regressor and DNN classifier	0.41 MAE locat. accur. 98.1%
CrossCount [117]	Human counting	WiFi/Passive/RSS	SISO/2.4/5GHz	RNN (LSTM)	Human count accur. 59%,
WiCount [94]	Human counting	WiFi/Passive/CSI	MIMO/5GHz	BP based DNN	Avg. counting accur. 82.3%
[159]	Crowd counting	WiFi/Passive/CSI	MIMO/5GHz	DNN	Crowd accur. 99% MCE 0.11-0.14
FreeCount [179]	Crowd counting	WiFi/Passive/CSI	MIMO/5GHz	Transfer kernel lear. and RBF based SVM	Avg. counting accur. 96%
DeMan [183]	Moving/ stationary individual detection	WiFi/Passive/CSI	MIMO/2.4GHz	SVM Classifier	Detection accur. 95%
Wi-CR [184]	Action counting and recognition	WiFi/Passive/CSI	MIMO/2.4GHz	KNN with DTW	Count. accur. 95% recog. accur. 90%
Wihi [99]	Gaits features	WiFi/Passive/CSI	MIMO/5GHz	RNN (LSTM)	Identifi. accur. 96/91% in group size of 2/8
GaitFi [118]	Gaits features	WiFi/Passive/CSI	MIMO/5GHz	Lightweight Residual convolution network	Recog. accur. 94.2%

understand the connections between the dynamics of CSI and the person's motion characteristics. The results show a remarkable performance, achieving accuracy at the decimeter level, with median location errors of approximately 24 cm (with initial positions) and 36 cm (without initial positions) along with an average relative velocity error of 11%.

The authors in [185] introduce WiTraj, an innovative device-less WiFi-based system for accurate indoor motion tracking. It enhances Doppler-frequency-shift (DFS) estimation by employing multiple angles to capture trajectory accurately, distinguish between walking and stationary actions, and improve tracking accuracy compared to existing methods. Experimental results show a median tracking error of less than 2.5% for room-sized trajectories. The study in [183] introduces DeMan, an innovative method for non-invasively detecting stationary and moving individuals using standard WiFi devices. DeMan employs amplitude and phase information from CSI for motion detection and utilizes subtle chest movements due to human respiration to identify stationary presence. DeMan achieves a remarkable detection rate of around 95% for both stationary and moving individuals with a 96% accuracy in identifying human-free scenarios and outperforms the existing benchmarks by approximately 30%. Similarly, the authors in [184] introduce Wi-CR which is a CSI-based technique for continuous action counting and recognition without requiring specialized hardware. Wi-CR improves action detection accuracy by identifying action start and end points, utilizing a peak-finding algorithm for counting consecutive actions. Two actions, walking and squats, are identified and distinguished. It integrates DWT-based waveform feature extraction and applies DTW and KNN for action recognition. Experimental results demonstrate that Wi-CR achieves 95% accuracy in action counting and 90% in action recognition. The authors in [186] developed a cost-effective and device-free approach that intelligently monitors human dynamics by estimating the number of participants, human density, and walking speed in a particular direction. The proposed technique employs the statistical distribution of CSI measurements to estimate the number of participants and density. At the same time, walking speed and direction are determined using a frequency-based mechanism. The experimental results demonstrate that the proposed system achieves over 90% accuracy in fine-grained human dynamics tracking – including participant count, density, walking speed, and direction across different indoor environments. Table 4 presents an overview of the existing works on tracking activities as a function of data sources, antenna systems, learning models, and key performance indicators (KPIs).

2) INDOOR LOCALIZATION

The performance of global positioning system (GPS) is not satisfactory in the indoor environment and WiFi fingerprinting can be utilized due to the widespread availability of WLAN in the indoor setup. However, the precision of indoor positioning is hampered by issues of uncertain and unstable fingerprints for the WiFi signals. To this end, the authors in [132] introduces FSTNet which is a DL framework to enhance indoor positioning accuracy by understanding the spatial-temporal relationship in the fingerprint data. The framework introduces the concept of path fingerprints to address uncertainty and instability in the fingerprint. It employs a CNN to effectively capture local attributes within path fingerprints and a fingerprint attention mechanism to efficiently capture spatial characteristics for stable positioning data. The on-site experiments on FSTNet demonstrates its ability to effectively capture temporal and spatial correlations within RSS measurements and achieves 44% enhancement in mean positioning error where 99.2% of errors being confined within the range of 2m. The indoor localization mainly relies on RSS and time of arrival (ToA) for fingerprinting. The study in [187] compared the three localization methods, i.e., multilateration, KNN, and minimum mean square error (MMSE) using RSS in the indoor WiFi systems. The aim is to assess their accuracy through simulations by correlating the estimated positions with actual data, and revealing the strengths and limitations of each approach via the network simulator (NS-2). The authors in [133] introduce a WiFi-based indoor localization system that leverages CNN for classification tasks. Three distinct approaches are explored such a custom architecture called WiFiNet, tailored for the task, and the use of popular pre-trained networks with transfer learning and feature extraction. Results shows that effectiveness of WiFiNet in indoor localization for medium-sized environments (30 positions, 113 access points) and reduce the localization error by 33% with less processing time compared to existing methods such as SVM.

The authors in [188] assess different learning models for indoor positioning within an office setting by utilizing an open source wireless infrastructure. The accuracy of estimated two-dimensional (2D) positions is evaluated in terms of root-mean-square error (RMSE) and R-squared metrics. The methods include distance estimation via an RSSI-based ML model, coupled with the Min-Max positioning algorithm, independent RSSI-based models for individual coordinates, and RSSI-based sequential prediction of the varying coordinates. A comparative analysis is conducted and the results reveal the superior performance of the distance prediction model and Min-Max algorithm for fixed positions and suggest the need for more precise sub-0.5 m measurements in 2D coordinate prediction. In [97], the authors proposed two DL models, i.e., a convolutional mixture density recurrent neural network and a variational autoencoder-based semi-supervised model for accurate user location estimation using limited labeled WiFi data. Validation experiments on real-world datasets confirm their superior performance compared to existing methods. The work in [189] presents a two-phase semi-supervised localization method applicable to indoor localization datasets. In the offline phase, the rank-based iterative clustering (RBIC) algorithm is used to form clusters with minimal distant points. RBIC acts as a clustering ensemble generated based on ranked baseline algorithms with clustering scores. In the online phase, user locations are estimated via classifiers using dynamic signal strength data. The system achieves high accuracy on the three benchmark datasets, i.e., 94%-99% for JUIndoorLoc, 96%-99%, and 95%-98% for the second and third datasets, respectively.

ViFi is introduced in [190] which is an indoor positioning system that merges the multi-wall multi-floor (MWMF) model for virtual reference points with a deterministic weighted k-Nearest Neighbors (WkNN) algorithm. ViFi outperforms prior virtual fingerprinting methods by maintaining high accuracy while reducing measurement time. Guidelines for parameter selection are provided and future research potential are discussed. The study in [98] presents a robust approach for tracking and locating moving objects in indoor the environments by using different classifier along with Kalman filtering and smoothing. The Feko channel-propagation emulator is used to simulate the RSSI maps in the indoor setting divided into multiple zones to enable multi-classification. A gradient-boosting decision-tree classifier is trained on RSSIs while Kalman filter enables online tracking and Kalman smoother facilitates offline localization. Results demonstrate 73.1% highest localization accuracy and the smallest average Euclidean-distance error of 1.33 m. Similarly, FedPos is presented in [134] which utilizes a federated transfer learning for WiFi CSI-based indoor localization. The overall improvement of 5.22% in the average accuracy is observed while the training time is reduced by about 34.78% compared to traditional learning methods. The authors in [149] evaluated the stacked autoencoder (SAE) for for three RSSI based indoor localization datasets and achieved an accuracy of 100% for building, 99.66% for the floor, and 83.47% for zone location. The RS-DeepNet is proposed in [125] for indoor localization with MAE of 4.81 m and 1.68m for two different indoor scenarios. A deep supervised autoencoder (SAE) based DeepPos is presented in [148] which achieve 1.9m improvement in MAE compared to DeepFi proposed in [191]. Similarly, a transformer-based indoor positioning system (TIPS) is presented in [153] that utilizes self-attention on WiFi CSI and direction of arrival (DoA) information to achieve distance error of up to 20cm, outperforming state-of-the-art methods. Table 5 presents the state-of-the-art on DL based indoor localization activities with an emphasis on the experimental setups, type of learning models, and the used KPIs such as positioning error, MMSE, localization accuracy, median error, and Euclidean distance error.

3) GAITS RECOGNITION

Human gait recognition identifies individuals from a distance and has gained popularity with time. With increasing data volume, the focus has shifted from traditional ML to advanced DL based gate recognition. In this regard, a brief review of DL techniques on gait identification using CNNs, capsule networks, RNNs, autoencoders, deep belief networks, and generative adversarial networks (GANs) is presented in [192]. AutoFi is proposed in [93], which utilizes

TABLE 5. WiFi-based indoor localization.

System/ Refere..	Tracking Activity	Experimental Setup		Learning Models	KPI
		Source/Type/ Measurement.	Number of APs and Measurements		
FSTNet [132]	Indoor	WiFi/Passive/RSSI	5-APs and 64-RPs	CNN	99.2% positioning error within 2m
[187]	Indoor	WiFi/Passive/RSSI/ToA	3-APs and 16-RPs	KNN	Around 0.04 MMSE
WiFiNet [133]	Indoor	WiFi/Passive/RSSI	113 APs with 11cross11 image	CNN	Accuracy 91.89% RMSE 28cm
[188]	Indoor	WiFi/Passive/RSSI	3APs and 3000 measurements	Linear Regression SVM and Ensemble	RMSE 2.52
[97]	Indoor	WiFi/Passive/RSSI	520 APs and 21049 measurements	CMDRNN VAE-based Semi-supervised learning	Datasets 1 and 2 Avg. Accu. 89.1%, 87.9% Median error 15%, 25%
[189]	Indoor	WiFi/Passive/RSSI	172 APs and 25364 measurements	RBIC algorithm with SVM classifier	JUIndoorLoc Dataset Accuracy 94%-99%
[98]	Indoor	WiFi/Passive/RSSI	1AP and 28RPs	KNN, RNN, and LSTM	Loc Accuracy 73.1% Avg. Eucl-dis. error 1.33m
[190]	Indoor	WiFi/Passive/RSS	6AP and 72RPs	KNN, WKNN	Avg. pos. error: 3.4m at spatial RP density: 0.02
[134]	Indoor	WiFi/Passive/CSI	1AP and 7RPs	CNN	mean loc. error 42.18cm
[125]	Indoor	WiFi/Passive/RSSI	Env.1: 4 APs, Env.2: 5 APs, 250 samples	GRU	Mean absolute error Env.1: 4.81m Env.2: 1.68m

low-quality CSI samples and employ MLP model for human gait recognition. In [136], the CSI data is preprocessed and then passed through the CNN model to detect walking direction, achieving an accuracy of 92.9%, 95.1%, and 89% in three indoor environments and across different people. The authors in [100] present Vi-WiFi-Gate which using RNN along with an attention mechanism and provides a gait recognition accuracy of 94.6% on over 1000 CSI samples. The authors in [193] introduce GAITWAY, an innovative system that identifies an individual's gait through walls using wireless radios. GAITWAY passively tracks gait speed with standard WiFi transceivers and eliminates the need for devices or restricted walkways. It detects stable walking periods, extracts relevant speed features, and identifies a person's gait. Using a rich-scattering multipath model, GAITWAY captures gait speed from over 10 m behind the walls. Experiments with 5,000 gait instances and various subjects demonstrate its accuracy and achieve a median speed error of 0.12 m/s, stride length error of 3.36 cm, and strong recognition rates of up to 81.2%. This establishes GAITWAY as both reliable and practical for use. The use of commercial WiFi devices for human identification has gained attention in applications such as smart homes and intrusion detection. However, the existing methods are prone to indoor noise and hence limited accuracy. The authors in [99] introduces Wihi, a new approach that employs DWT for noise removal, extracting human walking patterns via statistical features, and utilizing an RNN-LSTM model for accurate identification. Wihi's prototype on WiFi devices shows superior performance over existing methods showcasing its potential for robust human identity identification in noisy indoor environments. Human gait, a crucial identifier, can now be captured from a distance using passive sensors and applications in security and identification. While conventional research relies on cameras and computer vision for gait recognition, these methods struggle in low-light conditions. The article [118] presents GaitFi which is a multimodal technique that combines WiFi signals and videos. GaitFi employs CSI from WiFi along with camera-recorded videos and employs a lightweight residual convolution network (LRCN) to enhance gait information. By integrating WiFi and visual features in a two-stream model while training with triplet and classification loss, GaitFi achieves impressive real-world results with an accuracy of 94.2% in identifying 12 subjects and outperforming the WiFi only or camera-based methods. The paper [194] presents WiFiU, an approach that utilizes commercial WiFi devices to capture intricate gait patterns for human recognition. By exploiting the unique gait signatures in WiFi's CSI, the method generates spectrograms akin to Doppler RADARs through signal processing. Auto-correlation on the torso reflection further refines the pattern characterization. Evaluated on a dataset of 2,800 gait instances from 50 individuals in a 50m^2 room, WiFiU achieves recognition accuracy of 79.28% (top-1), 89.52% (top-2), and 93.05% (top-3). WiWho is presented in [195] which is a WiFi-based framework for device-free person identification in small groups. Utilizing WiFi CSI, WiWho captures unique gait patterns to distinguish individuals. Analyzing step and walk data within CSI, WiWho achieves accurate identification without personal devices while requiring a short walk. Evaluated with 20 volunteers across different locations, WiWho achieves 92% and 80% identification accuracy for 2 and 6-people groups, respectively. The study underscores that a 2-3m walk is sufficient for successful identification to enable a person's recognition in smart spaces using WiFi.

C. POSE ESTIMATION

HPE plays a vital role in human-computer interaction and can be applied to applications such as virtual reality and exercise monitoring in a smart environment. The traditional approaches rely on vision-based techniques which is not applicable in NLoS scenarios. Therefore, RF-based methods become an appealing alternative for pose estimation in scenarios with no wearable sensors and LoS requirements. Table 6 provides a comprehensive review of DL based HPE shedding light on considered dimensions, experimental setups, utilized learning models, and KPIs such as the percentage of correct key points (PCK@α) with distance threshold α, AP, prediction accuracy, and average joint localization error. The PCK@5 in Table 6 measures the proportion of key points where the distance between the predicted key point and the true joint falls within 5% of the subject's size. In other words, it refers to the percentage of keypoints where the predicted location is within 5% of the subject's size from the ground truth keypoints. The authors in [138] introduced a real-time approach for estimating the 2D poses of multiple people in an image. The method employs part affinity fields (PAFs) to effectively encode the limbs and orientations of humans, facilitating precise key point association. Feature extraction is accomplished using the first ten layers of visual geometry group (VGG-19) architecture which ensure favorable outcomes compared to the prior research by efficiently detecting the joints and associating connections among various body parts. Similarly, the authors in [135] introduce a novel approach for generating human pose images using wireless signals rather than optical cameras. The proposed framework combines data from multiple wireless devices capturing WiFi signals with an initial optical image from a camera. The wireless and vision data are then preprocessed before being fed into a CNN plus Residual network for generating human pose images. Experimental results indicate that this approach outperforms existing WiFi-based methods in terms of pose estimation accuracy and produces higher-quality visuals.

The paper [102] investigates the performance of WiFi-based HPE for single-person scenarios. An experiment is conducted using a setup comprising a 3-antenna WiFi transceiver to collect WiFi data. Simultaneously, a synchronized camera records videos of individuals with annotated key points are used for reference and a WiFi single-person pose network (WiSPPN) is proposed which is a fully connected CNN, designed to estimate single-person pose based on the collected CSI data and its corresponding annotations. The results demonstrate that WiSPPN achieves comparable accuracy to camera-based methods for single HPE. Similarly, the authors in [77] introduce a novel approach using the teacher-student network model for 2D human poses estimation in challenging scenarios involving walls and occlusion. The method utilizes synchronized visual data and signal reflections to address these difficulties. The teacher network [138] plays a vital role in the proposed approach with the employment of CNN with the ReLu activation function. Notably, despite not being explicitly trained for wall-related scenarios, the proposed model demonstrates impressive capabilities in estimating 2D human poses in such conditions. It achieves an acceptable average precision (AP) of 85.0%, which comparable with the teacher network having AP of 93.3%. The work in [103] proposes WiSPE, a 2D static HPE system based on commercial of-the-shelf (COTS) WiFi and utilizes 2D AoA imaging, resembling a camera, to estimate human pose accurately. The system incorporates the environment background filter (Env-Filter) algorithm to mitigate static environmental factors in the images and employs a teacher-student network (RNN and openpose) to correlate 2D AoA images with human skeleton joints. The results showed an average prediction accuracy of 95.2% in PCK@50 for each skeleton joint, surpassing other benchmarks and achieve a prediction accuracy of 73.1% in PCK@10.

The evolving attentive spatial-frequency network (EASFN) is introduced in [140], which combines static spatial and dynamic frequency information from dilated CSI sequences and utilize it for 2D HPE. The model includes an evolving attention module to focus on specific features and leads to a significant performance improvement of 16% in PCK @20 over the state-of-the-art method and proved to be effective for accurate HPE. The [160] study presents a multimodal HPE network (MHPEN) which is a unified network for 2D HPE, combining a teacher network (Alphapose) using visual data and a student network (PerUnet) utilizing WiFi CSI data. The WiFi-based HPE is focused by PerUnet which leverages the power of a multi-head attention mechanism and Unet-like architecture is used to effectively combine fine-grained pose features and contextual information from WiFi CSI. This integration leads to highly accurate HPE results. Additionally, the authors propose the attention-based denoising (ABD) method which effectively overcomes the limitations of traditional filters and facilitates the extraction of pose features from the CSI data. Extensive experiments are conducted, demonstrating that

TABLE 6. Pose estimation literature.

System/ Ref.	Pose Detail Dimensions /Keypoints/Poses	Experimental Setup		DL Model	KPI
		Source/Type/ Measure.	Antenna Sys. /Frequency		
WiSPPN [102]	2D/18-keypoints	WiFi/Passive/CSI	MIMO/2.4GHz	ResNet-based Enc/Dec	Avg. PCK@5 = 0.04 Avg. PCK @50 = 0.82
RF-Pose [77]	2D/14-keypoints/real-world	FMCW/Active /Heatmaps	MIMO/ (5.46 -7.24)GHz	CNN-based Enc/Dec	LOS AP = 62.4 NLOS AP = 58.1
Wispe [103]	2D AoA images/stand, sit, arm flat, arm up, akimbo	WiFi/Passive/AoA	SIMO/5GHz	RNN	Avg. prediction accur. at PCK@50 = 95.2% at PCK@10 = 73.1%
EASFN [140]	2D/17-keypoints	WiFi/Passive/CSI	MIMO/ISM-Band	CNN-based Enc/Dec	PCK@20 for SPE/ GPE = 50.05%/43.98%
PerUnet [160]	2D/18-keypoints/wave, walk, run, push, pull, jump, crouch, circle, sit down, stand up, throw	WiFi/Passive/CSI	MIMO /5GHz	UNet Enc/Dec	Avg. PCK@5 = 0.63 Avg. PCK@50 = 0.88
CSI-Former [139]	2D/18-keypoints/bend, circle, crouch, jump, pull, push, run, sit down, stand up, throw, walk,	WiFi/Passive/CSI	MIMO /5GHz	CNN-based Enc/Dec	PCK@5 = 0.5505
MDPose [101]	2D/17-keypoints/walk, sit, stand up, turn around, hit, covering, pick up, body rotation	WiFi-RADAR/ Passive/Micro-DS	MISO /ISM	CNN+ Bi/LSTM	Esti. mean abs err. D-based = 29.4mm M-based = 44mm
FastRFPose [161]	3D/walking poses	mmWave RADAR/ Active/Heatmap	MIMO/1.23GHz	HLN+PEN	HDA = 98.54% MPJPE = 11.05cm
WiPose [119]	3D/17-keypoints/lift hands, sweep hands, lift legs, wave hands, walk	WiFi/Passive/CSI	SIMO/5.825GHz	CNN+LSTM	Avg. joint loca. error for CSI = 30.1mm for 3DVP = 54.2mm
Gopose [120]	3D/14-keypoints/lift arms, wave hands, walk, jogging, use smartphone	WiFi/Passive/ CSI and AoAs	MIMO/5.32GHz	CNN+LSTM	Joint locali. err. range = (3.0-8.4)cm Avg. = 4.7cm
Wi-Mose [104]	3D/walk	WiFi/Passive/CSI	SIMO/5GHz	ResNet, RNN	P-MPJPE:LoS/NLOS = 29.7mm/37.8mm
CKD [142]	18-keypoints/stand, open arms, lie	WiFi + USRP /Passive/CSI	MIMO/5GHz	ResNet	Accur. improv. CKDloss = 0.3-0.1
Winect [143]	3D/17-keypoints/stand, open arms	WiFi/Passive/AoA	MIMO/5.32GHz	ResNet	Accur. = 98.5% AJLE = 4.6cm
Wi-Mesh [144], [145]	3D mesh/stand, walk	WiFi/Passive/AoA	MIMO/5.32GHz	ResNet	AVLE = 2.81cm JPE = 2.4cm
Wilink [146]	3D/free movements	WiFi/Passive/CSI	MIMO/5.2GHz	HPE(ResNet)	P-MPJPE = 32.31mm for train sub
DensePose [137]	UV maps	WiFi/Passive/CSI	MIMO/2.4GHz	MTN Enc/Dec ResNet-FPN	AP@75 to AP@50 = (35.6 - 87.2)

PerUnet achieves competitive performance in WiFi-based HPE evaluated on the Wi-Pose dataset. A novel architecture named CSI-former is introduced in [139] which integrates multi-head attention into a WiFi-based pose estimation network (PEN). The attention-based performer feature extraction (PAFE) and CNN-based encoder/decoder focus on information-rich CSI inputs which make the proposed method more promising over other HPE approaches. Experimental results on the Wi-Pose dataset demonstrate that CSI-former significantly enhances HPE performance with a PCK@5 accuracy of 0.5505 and surpasses the existing ResNet with PCK@5 accuracy of 0.5231. Similarly, a domain adaption algorithm called AdaPose is proposed in [196] for weakly supervised WiFi-based HPE which emphasizes consistent human poses to address environmental dynamics. in this regard, the CNN model and mapping consistency loss are utilized. The article [141] estimates head pose using CNN model WiFi CSI information including phase, amplitude, and frequency domain information such as phase, amplitude, and frequency domain information. The authors in [101] proposed an approach for reconstructing human skeletal motion through WiFi micro-Doppler signatures which enable practical human activity tracking with a 17-key point skeleton model. The method employs a CNN-RNN architecture to effectively learn temporal-spatial dependencies from clean micro-Doppler signatures. A pose optimization mechanism is applied to estimate the initial skeleton state and control error growth. Extensive testing in diverse environments with multiple subjects is performed using a single-receiver RADAR system. Furthermore, it is observed that the mean absolute error of 29.4mm is achieved for all key point positions. Similarly, MetaFi is presented for HPE in metaverse avatars simulation [197] which employs a customized DNN that is capable of achieving 95.23% accuracy at PCK@50 which is further improved to 97.30% in MetaFi++ [154].

A TGUL approach for 3D HPE is presented in [31] which employs a DNN with 9-residual block encoder for spatio-temporal convolutional features and a 3-residual block decoder for spatial de-convolution. Based on comparative analysis, predictive learning outperforms contrastive learning due to its ability to capture and utilize relevant information, resulting in consistent performance improvements over supervised baselines during fine-tuning. The authors in [161] introduce fast RFPose, a mmWave RADAR-based 3D HPE model which utilizes the human localization network (HLN) and the PEN network. HLN is responsible for predicting human positions in the RF heatmap and cropping the relevant regions while PEN estimates the 3D human poses based on these cropped areas. Extensive quantitative and qualitative analysis confirm the efficacy of Fast RFPose, demonstrating its ability to achieve accurate 3D HPE with remarkable processing speed. Furthermore, the trained Fast RFPose model is successfully deployed on a laptop with a central processing unit (CPU) underscoring its practical applicability in real-world settings. WiPose is an innovative 3D human pose construction framework that utilizes commercial WiFi devices to reconstruct accurate human skeletons in a challenging real-world scenario [119]. The DL model (CNN with LSTM) encodes prior knowledge of the human skeleton to ensure estimated joints adhere to the body's skeletal structure. Additionally, WiPose achieves cross-environment generalization by using a 3D velocity profile as input and effectively distinguishing posture-related features from static objects in the environment. Experimental results from real-world WiFi sensing testbed shows that WiPose achieves an average joint localization error of 2.83 cm and outperforms the state-of-the-art posture construction model designed for dedicated RADAR sensors by 35% accuracy. However, this method can only estimate the pose of a single human in a static position. GoPose is a 3D HPE system that utilizes WiFi signals from household devices [120]. By leveraging the 2D AoA spectrum of WiFi signals reflected off the human body and employing DL techniques such as CNN and LSTM models, the system achieves accurate pose estimation and tracking. Notably, it enables environment-independent pose estimation and exhibits an impressive accuracy of 4.7 cm across diverse scenarios including NLoS conditions and tracking unseen activities. This innovative technology holds significant potential for various applications requiring precise HPE. Wi-Mose utilizes the fusion of amplitude and phase of CSI data to create CSI images and enable the reconstruction of 3D poses of moving individuals while providing both pose and position information [104]. The system employs a residual network to extract pose-related features from these CSI images and a key-point regression network then converts these features into key-point coordinates. During training, synchronized visual data is used to supervise the WiFi data. The experimental results demonstrate Wi-Mose's effectiveness in accurately localizing key points and achieving mean-per-joint position error (P-MPJPE) values of 29.7 mm and 37.8 mm in LoS and NLoS scenarios, respectively. The authors in [147] introduce "Winect", which estimates 3D human poses from environment-independent free-form activities through the ResNet-18 model. The authors in [142] developed a correlated knowledge distillation (CKD) system using WiFi and USRP to detect human postures and movements while preserving privacy. CKD combined RF signals and camera images for a hybrid framework and used two parallel approaches (image and radio signal classification) for knowledge distillation. By leveraging correlated multimodal information between teacher and student networks, the CKD model achieved accurate inference without relying on images or video data. Experiments on the SDR-based testbed confirmed the feasibility and potential impact of this framework including the leverage of pre-trained networks with limited data. Similarly, the authors in [143] introduce Winect, an innovative 3D human pose tracking system that utilizes commodity WiFi devices to monitor 3D free-form activities. The method estimates a 3D skeleton pose with essential body joints to effectively track varied movements by integrating signal separation techniques and modeling joint motions. The system identifies moving limbs through the 2D AoA from human body reflections, untangles these signals for each limb, and constructs a comprehensive 3D skeleton by capturing the inherent correlation between limb motions and joint movements. The proposed system achieves an overall localization error of 4.6 cm, with joint localization errors ranging from 4.1 cm to 5.1 cm, showcasing Winect's ability to achieve cm-level precision in tracking diverse activities even in challenging scenarios, including non-line-of-sight situations. Similarly, [144], [145] propose Wi-Mesh which utilizes WiFi signals for creating a 3D human mesh. By capitalizing on the advancements in WiFi technology, the proposed system visualizes human body shapes and movements to construct a detailed 3D mesh. This is achieved by estimating the 2D AoA of WiFi signal reflections and essentially enables WiFi devices to perceive their surroundings. The process involves isolating human body images from the WiFi signal reflections and utilizing DL models to transform these images into a 3D mesh representation. Through rigorous experimentation across various indoor settings, Wi-Mesh demonstrates encouraging performance with an average error of 2.58 cm for vertices locations and 2.24 cm in joint positioning. WiLink is another WiFi-based 3D indoor HPE system that intelligently utilizes existing WiFi links for accurate HPE [146]. By classifying links as noise-dominated, most-effective, and redundant, the dynamic link selection (DLS) mechanism is introduced to adaptively choose the most-effective links by maximizing importance and minimizing redundancy. The selected CSI samples are fed into a DNN with residual blocks and fully connected layers for precise HPE. The P-MPJPE results (32.31 mm for training subjects and 40.30 mm for untrained subjects) demonstrate WiLink's cross-subject generalization. Similarly, the authors in [137] introduce a novel approach that utilizes WiFi signals and DL to estimate dense human pose correspondence. The DNN converts WiFi CSI into UV coordinates and is capable of achieving accurate pose estimations for 24 human regions. The proposed method incorporates a modality translation network and WiFi-DEnsePose RCNN to transform CSI into UV maps. Transfer Learning with a teacher network is used to further enhance the model performance. The results demonstrate that the WiFi-based model is more effective in estimating multiple subjects than the image-based methods. However, limitations exist, such as potential biases with rare body poses in the training data and challenges in handling multiple subjects in one capture. Nonetheless, the approach shows promising applications in HPE using readily available WiFi signals.

IV. CHALLENGES AND FUTURE DIRECTION

WiFi-based human sensing has received comprehensive attention in the existing literature and has demonstrated impressive accomplishments. On the other hand, DL is an effective tool that enhances the performance of WiFi sensing but still there exist enduring challenges that need to be addressed to fully exploit DL in future research. In this section, we present the challenges related to DL-based WiFi sensing and the potential strategic foresight.

A. PHYSICAL ENVIRONMENT VARIABILITY

In WiFi-based human sensing, there is a frequently underestimated but significant challenge related to environmental changes such as rearranging furniture, adding obstructions, human presence, closing doors, etc. Despite their seemingly common nature, these environmental alterations disrupt the consistency between the received data, like human activities, gestures, and locations, and the established reference data is used for sensing. For example, moving furniture or closing a door can unexpectedly change the behavior of WiFi signals and may cause unexpected reflections and distortion. As a result, these changes can lead to misunderstandings about human behavior and affect the overall performance of the sensing system [198]. To tackle this challenge, there is a need to create adaptive algorithms and DL techniques that can quickly adapt to these environmental shifts to ensure the reliability and robustness in real world dynamic situations.

B. MULTI-USER SENSING IN MULTI-MODAL SCENARIO

Most of the literature on WiFi sensing considers single-user scenarios. However, detecting and distinguishing the activities of multiple users simultaneously using commodity WiFi presents a significant challenge. This is particularly evident when trying to differentiate users based solely on the WiFi signal data since standard WiFi systems do not inherently provide user identification features. This challenge is further complicated by issues like signal interference and the blending of signals when several users are active at once. Additionally, it is crucial to prioritize data privacy as handling data collected from the activities of multiple users requires careful consideration to avoid violating their privacy rights. To advance this technology, sophisticated algorithms and signal processing techniques must be developed to accurately identify individual user activities while also preserving their privacy. This challenge remains complex and ongoing in the field of multi-user activity sensing using standard WiFi technology [199], [200]. On the other hand, the existence and availability of multiple technologies with different standards, principles, and operating frequencies is more ambitious in the era of future IoT. Most of the literature on sensing focuses on a single RF mode while the WiFi signals can be jointly utilized with other RF modes such as FMCW radar operating in higher frequency bands such as mm wave and terahertz, or even the sub-GHz LoRa signals [67]. It is an exciting future research direction to utilize and train DL on multi-stream data to enable multi-user sensing in multi-modal setups.

C. EFFICIENT DEEP LEARNING MODEL OR EDGE INTELLIGENCE FOR WIFI-BASED SENSING

Despite its strong sides, DL algorithms used in WiFi sensing face significant challenges. Firstly, the computational demands of complex DL models can hinder their real-time implementation on resource-constrained embedded devices, limiting their practicality in diverse scenarios. Secondly, gathering the extensive labeled datasets needed for training can be arduous and expensive, especially in situations where data collection is difficult. Additionally, DL algorithms often require substantial memory and storage resources, posing issues for devices with limited memory capacity. Furthermore, these models may lack transparency and interpretability, which is crucial in handling sensitive WiFi data. Addressing these limitations is vital for the widespread adoption of DL-based WiFi sensing systems, making the exploration of memory-efficient techniques and optimization methods a compelling avenue for research and progress in HAR [201].

D. HYBRID RF-VLC-BASED SENSING

Despite the advantage, there exist some challenges in RF-based systems such as spectrum congestion and interference issues. Therefore, high-frequency bands such as mmWaves are used for WiFi-based indoor localization and showed improved localization performance through spatial beam signal-to-noise ratios for RSSI fingerprinting [202]. Considering the energy consumption of RF-based systems, visible light offers several advantages such as spectrum availability, enhanced security, cost-effective implementation, and high energy efficiency [203]. Visible light communication (VLC) and localization have been extensively explored in literature [204], [205], [206]. However, despite these advantages, the visible light-based system requires LoS and results in limited performance in NLoS scenario [207]. Therefore, the heterogeneous deployment of WiFi technology and visible light can be a promising solution in the indoor environment as both are capable of improving the system performance by complementing each other [208]. Utilizing DL in such heterogeneous deployment for HAR, fall detection, HPE, motion sensing etc., can be one of the promising future research directions.

E. RECONFIGURABILITY AND MULTIPATH EFFECT FOR SMART INDOOR ENVIRONMENT

The multipath effect is a significant factor in WiFi-based human sensing that deserves close attention from researchers in the field. Multipaths occur when WiFi signals take multiple paths to reach a receiver influenced by reflections, diffraction, and scattering in indoor environments. In human sensing applications, this phenomenon presents notable challenges. As people move in a space, they can affect WiFi signal propagation by acting as signal reflectors. This can lead to unpredictable variations in signal strength and phase, potentially causing inaccuracies in detecting human presence and activities. Additionally, multipath interference can make it difficult to distinguish between multiple individuals nearby, affecting the system's ability to accurately track and identify users. To address the multipath effect in WiFi-based human sensing, innovative signal processing techniques, and advanced algorithms mitigate signal interference and improve system reliability. To enable a smart indoor environment, the reconfigurable intelligent surface (RIS) is considered a promising solution to improve the quality of the received signal by avoiding blockages, capacity, coverage, and energy efficiency [209]. However, RIS is mostly analyzed in the literature to improve the performance of wireless networks while there exist some works that consider the deployment of RIS for integrated sensing and communication through RF networks [210]. However, as mentioned in [211], the quality of received CSI is affected by RIS and can be considered to extend the capability of WiFi-based sensing in the indoor setup. Considering the joint use of DL and RIS in the smart indoor environment, we highlight the potential future research direction stated as follows.

1) DATASET GENERATION AND AVAILABILITY FOR RIS-ENABLED WIFI SENSING

In the context of WiFi-based sensing, creating standardized open datasets specifically designed for WiFi-based human sensing can make it easier and more cost-effective for researchers and developers to obtain labeled training data. Based on experimental data the performance of RIS-enabled WiFi sensing is an interesting future research direction. Keeping in view the importance of reproducible research, the dataset should be openly available so that new concepts and models in the area of ML are tested and compared with the existing ones.

2) AUTOMATED DATA LABELING IN RIS-ENABLED WIFI SENSING

Exploring automatic labeling methods using non-RF technologies, such as video or infrared sensors, offers a novel way to improve the efficiency and accuracy of data collection which is crucial in overcoming a key obstacle in developing robust WiFi sensing systems. Additionally, the concept of sensing with information fusion, which involves combining data from various sensors and sources, shows great potential for enhancing the reliability and precision of WiFi-based HAR. This approach can help address some of the challenges related to noisy signal data. While considering sensing with information fusion, using RIS as a sensing element is one of the promising future aspects of WiFi-based sensing. The DL model utilized in RIS-enabled WiFi sensing usually requires huge labeled data for training to learn and classify different features. However, automated data along with an efficient DL model needs to be designed to learn the unlabelled data to avoid the significant overhead in the data acquisition.

3) MULTIPLE RIS FOR MULTI-MODEL SENSING IN INDOOR WIFI SENSING

Utilizing RIS in the indoor system can improve the sensing capability of the RF system by utilizing improved CSI through the LoS link. As demonstrated in [212] RIS can play a vital role and its presence leads to achieving high accuracy in the recognition of human postures while in [213] a single RIS is proposed to enable a favorable propagation environment for RADAR sensing in the indoor environment. However, considering multiple path signals from multiple RIS in a multi-modal scenario can be interesting to explore in the future. Similarly, instead of single-user multi-user sensing capabilities while considering the impact of passive and active RIS needs to be analyzed. The location of RIS plays a vital role, therefore, optimal deployment of RIS can be explored for performance enhancement.

F. HITL BASED LEARNING MODEL FOR SENSING

A range of DL models are used for WiFi based human activity prediction. However, achieving optimal accuracy remains a primary goal for these models. One promising future research direction for WiFi-based human sensing is to improve the learning capabilities of DL models through Human-in-the-Loop" (HITL) methodologies. By integrating human feedback and interaction directly into the DL models enable them to adjust to various environments and user behaviors dynamically [214]. This approach capitalizes on real-time human inputs to correct and refine sensing outputs, thereby boosting system performance in complex real-world scenarios [215], [216]. For instance, users might validate or correct identified activities using a mobile application, allowing the system to learn from these interactions and update its models accordingly. This iterative process not only enhances model accuracy but also customizes the sensing experience to better fit individual user patterns and preferences. Moreover, the concept of HITL can be applied to various challenges in WiFi-based human sensing such as variability encountered in physical environments by empowering systems to adapt to changing conditions based on user feedback. In scenarios involving multiple users and modes, HITL can play a crucial role in managing and distinguishing signals from different users, enhancing the system's capability to handle complex environments. Additionally, the integration of HITL with DL or edge intelligence can optimize processing and reduce latency, making real-time adjustments more feasible. HITL also stands to improve reconfigurability and mitigate multipath effects in smart environments by leveraging user inputs to fine-tune system responses.

V. CONCLUSION

This study presents a comprehensive survey of recent research on human activity sensing via WiFi, leveraging DL algorithms. This study reveals the emerging trend of integrating DL with RF sensing has led to substantial progress. DL is a valuable technique to enhance the precision and range of device-free RF sensing. Researchers have effectively utilized DL to detect new phenomena that were previously not observed in the literature through traditional approaches. This survey briefly highlights the fundamental physical layer parameters used in RF sensing including RSSI, and CSI. Furthermore, the estimated parameters such as AoA and DS as well as FMCW technology followed by signal processing techniques is explored. Furthermore, we provided a detailed review on WiFi based human activity sensing such as motion detection, fall detection, crowd sensing, multi-activity classification, step sensing, HPE, localization, and gaits calculation assisted by DL techniques. Finally, the paper highlights the current limitations in WiFi-based sensing methods and presents the challenges that need to be addressed in the future. Based on the identified challenges, future research directions are suggested related to WiFi sensing to facilitate the reader's extending the research in the field.

REFERENCES

[1] F. Adib, Z. E. Kabelac, and D. Katabi, "Object tracking via radio reflections," U.S. Patent 15 120 864, Mar. 16, 2017.

[2] D. Katabi, "1.3 working at the intersection of machine learning, signal processing, sensors, and circuits," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2021, pp. 26–29.

[3] Y. He, Y. Chen, Y. Hu, and B. Zeng, "Wi-Fi vision: Sensing, recognition, and detection with commodity MIMO-OFDM Wi-Fi," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8296–8317, Sep. 2020.

[4] L. Lubna et al., "Radio frequency sensing and its innovative applications in diverse sectors: A comprehensive study," *Front. Commun. Netw.*, vol. 3, Sep. 2022, Art. no. 1010228.

[5] M. S. Mopas and E. Huybregts, "Training by feel: Wearable fitness-trackers, endurance athletes, and the sensing of data," *The Senses Soc.*, vol. 15, no. 1, pp. 25–40, 2020.

[6] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.

[7] K. C. Sahoo and U. C. Pati, "IoT based intrusion detection system using PIR sensor," in *Proc. IEEE 2nd Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, 2017, pp. 1641–1645.

[8] Q. Lin et al., "E-Jacket: Posture detection with loose-fitting garment using a novel strain sensor," in *Proc. 19th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Sydney, NSW, Australia, 2020, pp. 49–60.

[9] C. Zieger, A. Brutti, and P. Svaizer, "Acoustic based surveillance system for intrusion detection," in *Proc. IEEE 6th Int. Conf. Adv. Video Signal Based Surveillance*, Genova, Italy, 2009, pp. 314–319.

[10] Z. Shi, X. Chang, C. Yang, Z. Wu, and J. Wu, "An acoustic-based surveillance system for amateur drones detection and localization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2731–2739, Mar. 2020.

[11] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Las Vegas, NV, USA, 2016, pp. 426–429.

[12] F. Shahmohammadi, A. Hosseini, C. E. King, and M. Sarrafzadeh, "Smartwatch based activity recognition using active learning," in *Proc. IEEE/ACM Int. Conf. Connect. Health, Appl., Syst. Eng. Technol. (CHASE)*, Philadelphia, PA, USA, 2017, pp. 321–329.

[13] Q. Lei, J.-X. Du, H.-B. Zhang, S. Ye, and D.-S. Chen, "A survey of vision-based human action evaluation methods," *Sensors*, vol. 19, no. 19, p. 4129, 2019.

[14] G. Li and D. Wen, "Wearable biochemical sensors for human health monitoring: Sensing materials and manufacturing technologies," *J. Mater. Chem. B*, vol. 8, no. 16, pp. 3423–3436, 2020.

[15] A. Shastri et al., "A review of millimeter wave device-based localization and device-free sensing technologies and applications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1708–1749, 3rd Quart., 2022.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

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

[18] J. Yang, H. Zou, L. Xie, and C. J. Spanos, *Deep Learning and Unsupervised Domain Adaptation for Wi-Fi-Based Sensing*. Singapore: World Sci., 2021, ch. 4, pp. 79–100.

[19] C. Shi, J. Liu, N. Borodinov, B. Leao, and Y. Chen, "Towards environment-independent behavior-based user authentication using Wi-Fi," in *Proc. IEEE 17th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, 2020, pp. 666–674.

[20] S. Fang, T. Islam, S. Munir, and S. Nirjon, "EyeFi: Fast human identification through vision and Wi-Fi-based trajectory matching," in *Proc. 16th Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS)*, Marina del Rey, CA, USA, 2020, pp. 59–68.

[21] Y. Zhang, C. Wu, and Y. Chen, "A low-overhead indoor positioning system using CSI fingerprint based on transfer learning," *IEEE Sensors J.*, vol. 21, no. 16, pp. 18156–18165, Aug. 2021.

[22] S. A. Shah and F. Fioranelli, "RF sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 11, pp. 26–44, Nov. 2019.

[23] X. Wang, X. Wang, and S. Mao, "RF sensing in the Internet of Things: A general deep learning framework," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 62–67, Sep. 2018.

[24] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83791–83820, 2020.

[25] Z. Hussain, M. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: A survey," 2019, *arXiv:1906.05074*.

[26] Y. Yao et al., "Fall detection system using millimeter-wave radar based on neural network and information fusion," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21038–21050, Nov. 2022.

[27] 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 J.*, vol. 21, no. 12, pp. 13522–13529, Jun. 2021.

[28] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, Jul. 2019.

[29] X. Dong, Z. Zhao, Y. Wang, T. Zeng, J. Wang, and Y. Sui, "FMCW radar-based hand gesture recognition using spatiotemporal deformable and context-aware convolutional 5-D feature representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2021, Art. no. 5107011.

[30] A. Chebrolu, "FallWatch: A novel approach for through-wall fall detection in real-time for the elderly using artificial intelligence," in *Proc. IEEE 3rd Int. Conf. Transdiscipl. AI (TransAI)*, Laguna Hills, CA, USA, 2021, pp. 57–63.

[31] T. Li, L. Fan, Y. Yuan, and D. Katabi, "Unsupervised learning for human sensing using radio signals," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Waikoloa, HI, USA, 2022, pp. 3288–3297.

[32] L. Fan, T. Li, Y. Yuan, and D. Katabi, "In-home daily-life captioning using radio signals," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K., 2020, pp. 105–123.

[33] L. Fan, T. Li, R. Fang, R. Hristov, Y. Yuan, and D. Katabi, "Learning long-term representations for person re-identification using radio signals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2020, pp. 10699–10709.

[34] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling identification and behavioral sensing in homes using radio reflections," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Glasgow, U.K., 2019, pp. 1–13.

[35] C. Ding, Y. Zou, L. Sun, H. Hong, X. Zhu, and C. Li, "Fall detection with multi-domain features by a portable FMCW radar," in *Proc. IEEE MTT-S Int. Wireless Symp. (IWS)*, Guangzhou, China, 2019, pp. 1–3.

[36] Y. Zhao, V. Sark, M. Krstic, and E. Grass, "Novel approach for gesture recognition using mmwave FMCW radar," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC)*, Helsinki, Finland, 2022, pp. 1–6.

[37] C. Wang et al., "Multi-touch in the Air: Device-free finger tracking and gesture recognition via COTS RFID," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Honolulu, HI, USA, 2018, pp. 1691–1699.

[38] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2017.

[39] H. Ding et al., "Femo: A platform for free-weight exercise monitoring with RFIDs," in *Proc. 13th ACM Conf. Embed. Netw. Sensor Syst.*, Seoul, South Korea, 2015, pp. 141–154.

[40] Y. Wang and Y. Zheng, "Modeling RFID signal reflection for contact-free activity recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 1–22, 2018.

[41] G. A. Oguntala et al., "Smartwall: Novel RFID-enabled ambient human activity recognition using machine learning for unobtrusive health monitoring," *IEEE Access*, vol. 7, pp. 68022–68033, 2019.

[42] X. Fan, W. Gong, and J. Liu, "Tagfree activity identification with RFIDs," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–23, 2018.

[43] K. Toda and N. Shinomiya, "Machine learning-based fall detection system for the elderly using passive RFID sensor tags," in *Proc. IEEE 13th Int. Conf. Sens. Technol. (ICST)*, Sydney, NSW, Australia, 2019, pp. 1–6.

[44] X. Fan, F. Wang, F. Wang, W. Gong, and J. Liu, "When RFID meets deep learning: Exploring cognitive intelligence for activity identification," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 19–25, Jun. 2019.

[45] Z. Liu, X. Liu, J. Zhang, and K. Li, "Opportunities and challenges of wireless human sensing for the smart IoT world: A survey," *IEEE Netw.*, vol. 33, no. 5, pp. 104–110, Oct. 2019.

[46] Z. Wang, Z. Huang, C. Zhang, W. Dou, Y. Guo, and D. Chen, "CSI-based human sensing using model-based approaches: A survey," *J. Comput. Design Eng.*, vol. 8, no. 2, pp. 510–523, 2021.

[47] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[48] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity W-Fi," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Network.*, Paris, France, 2015, pp. 53–64.

[49] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger: Talk to your smart devices with finger-grained gesture," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, Heldelberg, Germany, 2016, pp. 250–261.

[50] D. Zhu, N. Pang, G. Li, and S. Liu, "WiseFi: Activity localization and recognition on commodity off-the-shelf Wi-Fi devices," in *Proc. IEEE 18th Int. Conf. High Perform. Comput. Commun.*, Sydney, NSW, Australia, 2016, pp. 562–569.

[51] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using Wi-Fi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[52] Z. Wang et al., "A survey on CSI-based human behavior recognition in through-the-wall scenario," *IEEE Access*, vol. 7, pp. 78772–78793, 2019.

[53] H. F. T. Ahmed, H. Ahmad, and C. Aravind, "Device free human gesture recognition using Wi-Fi CSI: A survey," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103281.

[54] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2019.

[55] J. Xiao, H. Li, M. Wu, H. Jin, M. J. Deen, and J. Cao, "A survey on wireless device-free human sensing: Application scenarios, current solutions, and open issues," *ACM Comput. Surveys*, vol. 55, no. 5, pp. 1–35, 2022.

[56] Y. Ge et al., "Contactless Wi-Fi sensing and monitoring for future healthcare-emerging trends, challenges, and opportunities," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 171–191, 2022.

[57] J. Yang et al., "SenseFi: A library and benchmark on deep-learning-empowered Wi-Fi human sensing," *Patterns*, vol. 4, no. 3, 2023, Art. no. 100703.

[58] J. Ding, Y. Wang, J. Zhang, H. Chen, H. Si, and J. Liu, "Multi-AP CSI fusion and features optimization-based behavioral sensing on Wi-Fi platform," *IEEE Sensors J.*, vol. 23, no. 22, pp. 27601–27612, Nov. 2023.

[59] M. Ali, P. Hendriks, N. Popping, S. Levi, and A. Naveed, "A comparison of machine learning algorithms for Wi-Fi sensing using CSI data," *Electronics*, vol. 12, no. 18, p. 3935, 2023.

[60] P. F. Moshiri, R. Shahbazian, M. Nabati, and S. A. Ghorashi, "A CSI-based human activity recognition using deep learning," *Sensors*, vol. 21, no. 21, p. 7225, 2021.

[61] F. Succetti et al., "A fast deep learning technique for Wi-Fi-based human activity recognition," *Electrom. Waves*, vol. 174, pp. 127–141, Jul. 2022.

[62] A. Poulose, J. H. Kim, and D. S. Han, "HIT HAR: Human image threshing machine for human activity recognition using deep learning models," *Comput. Intell. Neurosci.*, vol. 2022, Oct. 2022, Art. no. 1808990.

[63] M. Keaton et al., "Wi-Fi-based in-home fall-detection utility: Application of Wi-Fi channel state information as a fall detection service," in *Proc. IEEE Int. Conf. Eng., Technol. Innovat. (ICE/ITMC)*, Cardiff, U.K., 2020, pp. 1–6.

[64] T. Nakamura, M. Bouazizi, K. Yamamoto, and T. Ohtsuki, "Wi-Fi-based fall detection using spectrogram image of channel state information," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17220–17234, Sep. 2022.
[65] U. M. Khan, Z. Kabir, S. A. Hassan, and S. H. Ahmed, "A deep learning framework using passive Wi-Fi sensing for respiration monitoring," in *Proc. IEEE Global Commun. Conf. (GLOBCOM)*, Singapore, 2017, pp. 1–6.
[66] R. Zhang, X. Jing, S. Wu, C. Jiang, J. Mu, and F. R. Yu, "Device-free wireless sensing for human detection: The deep learning perspective," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2517–2539, Feb. 2021.
[67] J. Zhang et al., "A survey of mmwave-based human sensing: Technology, platforms and applications," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2052–2087, 4th Quart., 2023.
[68] S. Pisa, E. Pittella, and E. Piuzzi, "A survey of radar systems for medical applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 31, no. 11, pp. 64–81, Nov. 2016.
[69] E. Elbasani, P. Siriporn, and J. S. Choi, *A Survey on RFID in Industry 4.0*. Cham, Switzerland: Springer, 2020, pp. 1–16.
[70] Y. Duroc, "From identification to sensing: RFID is one of the key technologies in the IoT field," *Sensors*, vol. 22, no. 19, p. 7523, 2022.
[71] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu, "Smart home based on Wi-Fi sensing: A survey," *IEEE Access*, vol. 6, pp. 13317–13325, 2018.
[72] D. Wu, D. Zhang, C. Xu, H. Wang, and X. Li, "Device-free Wi-Fi human sensing: From pattern-based to model-based approaches," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 91–97, Oct. 2017.
[73] P. Bahl and V. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. 19th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, Tel Aviv, Israel, 2000, pp. 775–784.
[74] Y. Wang, M. Li, and M. Li, "The statistical analysis of IEEE 802.11 wireless local area network–based received signal strength indicator in indoor location sensing systems," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 12, 2017, Art. no. 1550147717747858.
[75] A. T. Parameswaran, M. I. Husain, and S. Upadhyaya, "Is RSSI a reliable parameter in sensor localization algorithms: An experimental study," in *Proc. IEEE Field Failure Data Analysis Workshop (F2DA09)*, Niagara Falls, NY, USA, 2009, pp. 1–5.
[76] R. Zavorka, R. Marsalek, J. Vychodil, E. Zöchmann, G. Ghiaasi, and J. Blumenstein, "Deep neural network-based human activity classifier in 60 GHz WLAN channels," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Rio de Janeiro, Brazil, 2022, pp. 1304–1309.
[77] M. Zhao et al., "Through-wall human pose estimation using radio signals," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 7356–7365.
[78] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Network.*, Los Cabos, Mexico, 2019, pp. 1–16.
[79] K. Chetty, G. E. Smith, and K. Woodbridge, "Through-the-wall sensing of personnel using passive bistatic Wi-Fi radar at standoff distances," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1218–1226, Apr. 2012.
[80] B. Tan, K. Woodbridge, and K. Chetty, "Awireless passive radar system for real-time through-wall movement detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2596–2603, Oct. 2016.
[81] B. Tan, K. Woodbridge, and K. Chetty, "A real-time high resolution passive Wi-Fi doppler-radar and its applications," in *Proc. Int. Radar Conf.*, Lille, France, 2014, pp. 1–6.
[82] M. AlHamaydeh and N. Ghazal Aswad, "Structural health monitoring techniques and technologies for large-scale structures: Challenges, limitations, and recommendations," *Pract. Periodical Struct. Design Constr.*, vol. 27, no. 3, 2022, Art. no. 3122004.
[83] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "mmASL: Environment-independent ASl gesture recognition using 60 GHz millimeter-wave signals," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–30, 2020.
[84] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of Wi-Fi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, 2015, pp. 65–76.
[85] S. Li, Z. Liu, Y. Zhang, Q. Lv, X. Niu, L. Wang, and D. Zhang, "WiBorder: Precise Wi-Fi based boundary sensing via through-wall discrimination," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–30, 2020.
[86] Y. Ma, G. Zhou, and S. Wang, "Wi-Fi sensing with channel state information: A survey," *ACM Comput. Surveys*, vol. 52, no. 3, pp. 1–36, 2019.
[87] J. Shang and J. Wu, "Fine-grained vital signs estimation using commercial Wi-Fi devices," in *Proc. 8th Wireless Students, Workshop*, New York, NY, USA, 2016, pp. 30–32.
[88] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous Wi-Fi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, China, 2015, pp. 1472–1480.
[89] N. Xiao, P. Yang, Y. Yan, H. Zhou, and X.-Y. Li, "Motion-Fi: Recognizing and counting repetitive motions with passive wireless backscattering," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Honolulu, HI, USA, 2018, pp. 2024–2032.
[90] T. Yu, H. Jin, and K. Nahrstedt, "Writinghacker: Audio-based eavesdropping of handwriting via mobile devices," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 463–473.
[91] M. N. Patil, B. Iyer, and R. Arya, "Performance evaluation of PCA and ICA algorithm for facial expression recognition application," in *Proc. 5th Int. Conf. Soft Comput. Problem Solving*, Singapore, 2016, pp. 965–976.
[92] Z. Yang, Y. Zhang, and Q. Zhang, "Rethinking fall detection with Wi-Fi," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 6126–6143, Oct. 2023.
[93] J. Yang, X. Chen, H. Zou, D. Wang, and L. Xie, "AutoFi: Toward automatic Wi-Fi human sensing via geometric self-supervised learning," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 7416–7425, Apr. 2023.
[94] S. Liu, Y. Zhao, and B. Chen, "WiCount: A deep learning approach for crowd counting using Wi-Fi signals," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl. IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC)*, Guangzhou, China, 2017, pp. 967–974.
[95] Y. Ma et al., "Location-and person-independent activity recognition with Wi-Fi, deep neural networks, and reinforcement learning," *ACM Trans. Internet Things*, vol. 2, no. 1, pp. 1–25, 2021.
[96] L. Hao, B. Huang, B. Jia, G. Xu, and G. Mao, "Toward accurate crowd counting in large surveillance areas based on passive Wi-Fi sensing," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 14086–14096, Dec. 2023.
[97] W. Qian, F. Lauri, and F. Gechter, "Supervised and semi-supervised deep probabilistic models for indoor positioning problems," *Neurocomputing*, vol. 435, pp. 228–238, May 2021.
[98] G. Liu, P. Neupane, H.-C. Wu, W. Xiang, L. Pu, and S. Y. Chang, "Novel robust indoor device-free moving-object localization and tracking using machine learning with Kalman filter and smoother," *IEEE Syst. J.*, vol. 16, no. 4, pp. 6253–6264, Dec. 2022.
[99] J. Ding, Y. Wang, and X. Fu, "Wihi: Wi-Fi based human identity identification using deep learning," *IEEE Access*, vol. 8, pp. 129246–129262, 2020.
[100] T. Li, T. Chen, K. Li, and R. Ao, "Vi-Wi-Fi-Gate: A Wi-Fi sensing gait recognition method inspired by the image vision field," in *Proc. Int. Conf. Internet Things Mach. Learn. (IoTML)*, Singapore, 2023, pp. 376–381.
[101] C. Tang, W. Li, S. Vishwakarma, F. Shi, S. Julier, and K. Chetty, "MDPose: Human skeletal motion reconstruction using Wi-Fi micro-doppler signatures," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 1, pp. 157–167, Feb. 2024.
[102] F. Wang, S. Panev, Z. Dai, J. Han, and D. Huang, "Can Wi-Fi estimate person pose?" 2019, *arXiv:1904.00277*.
[103] M. Xu, Z. Guo, L. Gui, B. Sheng, and F. Xiao, "WiSPE: A COTS Wi-Fi-based 2D static human pose estimation," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3560–3571, Sep. 2023.
[104] Y. Wang, L. Guo, Z. Lu, X. Wen, S. Zhou, and W. Meng, "From point to space: 3D moving human pose estimation using commodity Wi-Fi," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2235–2239, Jul. 2021.
[105] T.-D. H. Nguyen and H.-N. H. Nguyen, "Towards a robust Wi-Fi-based fall detection with adversarial data augmentation," in *Proc. IEEE 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, 2020, pp. 1–6.

[106] R. Guo, H. Li, D. Han, and R. Liu, "Feasibility analysis of using channel state information (CSI) acquired from Wi-Fi routers for construction worker fall detection," *Int. J. Environ. Res. Public Health*, vol. 20, no. 6, p. 4998, 2023.

[107] F. Wang, W. Gong, and J. Liu, "On spatial diversity in Wi-Fi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2035–2047, Apr. 2019.

[108] W. Cui, B. Li, L. Zhang, and Z. Chen, "Device-free single-user activity recognition using diversified deep ensemble learning," *Appl. Soft Comput.*, vol. 102, Apr. 2021, Art. no. 107066.

[109] F. Abuhoureyah, Y. C. Wong, A. S. B. M. Isira, and M. N. Al-Andoli, "CSI-based location independent human activity recognition using deep learning," *Human-Centric Intell. Syst.*, vol. 3, pp. 1–21, Dec. 2023.

[110] A. Elkelany, R. Ross, and S. Mckeever, "Wi-Fi-based human activity recognition using attention-based BiLSTM," in *Proc. Irish Conf. Artif. Intell. Cogn. Sci.*, Munster, Ireland, 2022, pp. 121–133.

[111] M. J. Bocus, K. Chetty, and R. J. Piechocki, "UWB and Wi-Fi systems as passive opportunistic activity sensing radars," in *Proc. IEEE Radar Conf. (RadarConf21)*, Atlanta, GA, USA, 2021, pp. 1–6.

[112] A. Zhuravchak, O. Kapshii, and E. Pournaras, "Human activity recognition based on Wi-Fi CSI data-a deep neural network approach," *Procedia Comput. Sci.*, vol. 198, pp. 59–66, Jan. 2022.

[113] Y. Zhang, Y. Yin, Y. Wang, J. Ai, and D. Wu, "CSI-based location-independent human activity recognition with parallel convolutional networks," *Comput. Commun.*, vol. 197, pp. 87–95, Jan. 2023.

[114] S. Mekruksavanich, W. Phaphan, N. Hnoohom, and A. Jitpattanakul, "Attention-based hybrid deep learning network for human activity recognition using Wi-Fi channel state information," *Appl. Sci.*, vol. 13, no. 15, p. 8884, 2023.

[115] L. Hao, B. Huang, B. Jia, and G. Mao, "On the fine-grained crowd analysis via passive Wi-Fi sensing," *IEEE Trans. Mobile Comput.*, vol. 23, no. 6, pp. 6697–6711, Jun. 2024.

[116] U. Singh, J.-F. Determe, F. Horlin, and P. De Doncker, "Crowd forecasting based on Wi-Fi sensors and LSTM neural networks," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6121–6131, Sep. 2020.

[117] O. T. Ibrahim, W. Gomaa, and M. Youssef, "CrossCount: A deep learning system for device-free human counting using Wi-Fi," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9921–9928, Nov. 2019.

[118] L. Deng, J. Yang, S. Yuan, H. Zou, C. X. Lu, and L. Xie, "GaitFi: Robust device-free human identification via Wi-Fi and vision multimodal learning," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 625–636, Jan. 2023.

[119] W. Jiang et al., "Towards 3D human pose construction using Wi-Fi," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, London, U.K., 2020, pp. 1–14.

[120] Y. Ren, Z. Wang, Y. Wang, S. Tan, Y. Chen, and J. Yang, "GoPose: 3D human pose estimation using Wi-Fi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 1–25, 2022.

[121] J. Liu, L. Wang, L. Guo, J. Fang, B. Lu, and W. Zhou, "A research on CSI-based human motion detection in complex scenarios," in *Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Dalian, China, 2017, pp. 1–6.

[122] Z. He, X. Zhang, Y. Wang, Y. Lin, G. Gui, and H. Gacanin, "A robust CSI-based Wi-Fi passive sensing method using attention mechanism deep learning," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 17490–17499, Oct. 2023.

[123] S. M. Bokhari, S. Sohaib, A. R. Khan, and M. Shafi, "DGRU based human activity recognition using channel state information," *Measurement*, vol. 167, Jan. 2021, Art. no. 108245.

[124] L. Zhang, Y. Zhang, R. Bao, Y. Zhu, and X. Shi, "A novel Wi-Fi-based personnel behavior sensing with a deep learning method," *IEEE Access*, vol. 10, pp. 120136–120145, 2022.

[125] F. Khan, A. U. Khan, I. Ahmad, and W. Choi, "RS-DeepNet: A machine learning aided RSSI fingerprinting for precise indoor localization," in *Proc. 2nd Int. Conf. Maritime IT Converg. (ICMIC)*, Jeju, South Korea, 2023, pp. 1–6.

[126] H. El Zein, F. Mourad-Chehade, and H. Amoud, "Leveraging Wi-Fi CSI data for fall detection: A deep learning approach," in *Proc. IEEE 5th Int. Conf. Bio-Eng. Smart Technol. (BioSMART)*, Paris, France, 2023, pp. 1–4.

[127] C. Xiao, Y. Lei, Y. Ma, F. Zhou, and Z. Qin, "DeepSeg: Deep-learning-based activity segmentation framework for activity recognition using Wi-Fi," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5669–5681, Apr. 2021.

[128] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, 2019.

[129] Y. Wang, X. Jiang, R. Cao, and X. Wang, "Robust indoor human activity recognition using wireless signals," *Sensors*, vol. 15, no. 7, pp. 17195–17208, 2015.

[130] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "Rt-fall: A real-time and contactless fall detection system with commodity Wi-Fi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[131] C. Zhang and W. Jiao, "Imgfi: A high accuracy and lightweight human activity recognition framework using CSI image," *IEEE Sensors J.*, vol. 23, no. 18, pp. 21966–21977, Sep. 2023.

[132] Q. Ren, Y. Wang, S. Liu, and X. Lv, "FSTNet: Learning spatial-temporal correlations from fingerprints for indoor positioning," *Ad Hoc Netw.*, vol. 149, Oct. 2023, Art. no. 103244.

[133] N. Hernández et al., "Wi-FiNet: Wi-Fi-based indoor localisation using CNNs," *Expert Syst. Appl.*, vol. 177, Sep. 2021, Art. no. 114906.

[134] J. Guo, I. W.-H. Ho, Y. Hou, and Z. Li, "FedPos: A federated transfer learning framework for CSI-based Wi-Fi indoor positioning," *IEEE Syst. J.*, vol. 17, no. 3, pp. 4579–4590, Sep. 2023.

[135] C. Yu et al., "Wi-Fi-based human pose image generation," in *Proc. IEEE 24th Int. Workshop Multimedia Signal Process. (MMSP)*, Shanghai, China, 2022, pp. 1–6.

[136] H. A. H. Ali and S. Seytnazarov, "Human walking direction detection using wireless signals, machine and deep learning algorithms," *Sensors*, vol. 23, no. 24, p. 9726, 2023.

[137] J. Geng, D. Huang, and F. De la Torre, "DensePose from Wi-Fi," 2022, *arXiv:2301.00250*.

[138] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 7291–7299.

[139] Y. Zhou, C. Xu, L. Zhao, A. Zhu, F. Hu, and Y. Li, "Wi-Fi-Former: Pay more attention to pose estimation with Wi-Fi," *Entropy*, vol. 25, no. 1, p. 20, 2022.

[140] Y.-C. Chen et al., "Seeing the unseen: Wi-Fi-based 2D human pose estimation via an evolving attentive spatial-frequency network," *Pattern Recognit. Lett.*, vol. 171, pp. 21–27, Jul. 2023.

[141] Y. Liu and S. Konomi, "Wihead: Wi-Fi-based head-pose estimation," in *Proc. Int. Conf. Human-Comput. Interact.*, Cham, Switzerland, 2022, pp. 69–86.

[142] S. R. Pokhrel et al., "On correlated knowledge distillation for monitoring human pose with radios," 2023, *arXiv:2305.14829*.

[143] Y. Ren, Z. Wang, S. Tan, Y. Chen, and J. Yang, "Winect: 3D human pose tracking for free-form activity using commodity Wi-Fi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 4, pp. 1–29, 2021.

[144] Y. Wang, Y. Ren, Y. Chen, and J. Yang, "A Wi-Fi vision-based 3D human mesh reconstruction," in *Proc. 28th Annu. Int. Conf. Mobile Comput. Netw.*, Sydney, NSW, Australia, 2022, pp. 814–816.

[145] Y. Wang and J. Yang, "3D human mesh construction leveraging Wi-Fi," 2022, *arXiv:2210.10957*.

[146] L. Wang, L. Guo, Z. Lu, X. Wen, and S. Zhou, "WiLink: Link selection-based 3D human pose estimation using commodity Wi-Fi," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Glasgow, U.K., 2023, pp. 1–7.

[147] Y. Ren and J. Yang, "3D human pose estimation for free-form activity using Wi-Fi signals," 2021, *arXiv:2110.08314*.

[148] P. Yazdanian and V. Pourahmadi, "DeepPos: Deep supervised autoencoder network for CSI-based indoor localization," 2018, *arXiv:1811.12182*.

[149] S. BelMannoubi, H. Touati, and H. Snoussi, "Stacked auto-encoder for scalable indoor localization in wireless sensor networks," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Tangier, Morocco, 2019, pp. 1245–1250.

[150] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 286–293.

[151] F. Luo, S. Khan, B. Jiang, and K. Wu, "Vision transformers for human activity recognition using Wi-Fi channel state information," *IEEE Internet Things J.*, early access, Mar. 21, 2024, doi: 10.1109/JIOT.2024.3375337.
[152] B. Sheng, R. Han, H. Cai, F. Xiao, L. Gui, and Z. Guo, "CDFi: Cross-domain action recognition using Wi-Fi signals," *IEEE Trans. Mobile Comput.*, early access, Jan. 2, 2024, doi: 10.1109/TMC.2023.3348939.
[153] Z. Zhang, H. Du, S. Choi, and S. H. Cho, "TIPS: Transformer based indoor positioning system using both CSI and DoA of Wi-Fi signal," *IEEE Access*, vol. 10, pp. 111363–111376, 2022.
[154] Y. Zhou, H. Huang, S. Yuan, H. Zou, L. Xie, and J. Yang, "MetaFi++: Wi-Fi-enabled transformer-based human pose estimation for metaverse avatar simulation," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14128–14136, Aug. 2023.
[155] Y. Chu, K. Cumanan, S. K. Sankarpandi, S. Smith, and O. A. Dobre, "Deep learning-based fall detection using Wi-Fi channel state information," *IEEE Access*, vol. 11, pp. 83763–83780, 2023.
[156] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, Long Beach, California, USA, 2019, pp. 6105–6114.
[157] H. Choi, T. Matsui, S. Misaki, A. Miyaji, M. Fujimoto, and K. Yasumoto, "Simultaneous crowd estimation in counting and localization using Wi-Fi CSI," in *Proc. IEEE Int. Conf. Indoor Position. Indoor Navig. (IPIN)*, Lloret de Mar, Spain, 2021, pp. 1–8.
[158] H. Choi, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-Cal: Wi-Fi sensing and machine learning based device-free crowd counting and localization," *IEEE Access*, vol. 10, pp. 24395–24410, 2022.
[159] R. Zhou, X. Lu, Y. Fu, and M. Tang, "Device-free crowd counting with Wi-Fi channel state information and deep neural networks," *Wireless Netw.*, vol. 26, pp. 3495–3506, Jul. 2020.
[160] Y. Zhou, A. Zhu, C. Xu, F. Hu, and Y. Li, "PerUnet: Deep signal channel attention in Unet for Wi-Fi-based human pose estimation," *IEEE Sensors J.*, vol. 22, no. 20, pp. 19750–19760, Oct. 2022.
[161] C. Yu et al., "Fast 3D human pose estimation using RF signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, 2023, pp. 1–5.
[162] S. Pouyanfar et al., "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surveys*, vol. 51, no. 5, pp. 1–36, 2018.
[163] J. Ding and Y. Wang, "A Wi-Fi-based smart home fall detection system using recurrent neural network," *IEEE Trans. Consum. Electron.*, vol. 66, no. 4, pp. 308–317, Nov. 2020.
[164] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Boston, MA, USA, 2017, pp. 1597–1600.
[165] W. Wang, Y. Huang, Y. Wang, and L. Wang, "Generalized autoencoder: A neural network framework for dimensionality reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Columbus, OH, USA, 2014, pp. 490–497.
[166] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.
[167] S. Chen, W. Yang, Y. Xu, Y. Geng, B. Xin, and L. Huang, "AFall: Wi-Fi-based device-free fall detection system using spatial angle of arrival," *IEEE Trans. Mobile Comput.*, vol. 22, no. 8, pp. 4471–4484, Aug. 2023.
[168] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wi-Fi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.
[169] W. Jiang et al., "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 289–304.
[170] Z. Shi, Q. Cheng, J. A. Zhang, and R. Y. Da Xu, "Environment-robust Wi-Fi-based human activity recognition using enhanced CSI and deep learning," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 24643–24654, Dec. 2022.
[171] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Trans. Mobile Comput.*, vol. 21, no. 2, pp. 540–554, Feb. 2022.
[172] "Falls." WHO. 2021. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/falls
[173] "Ageing and health." WHO. 2022. [Online]. Available: https://www.who.int/ageing/about/facts/en/.
[174] Y. Hu, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "A Wi-Fi-based passive fall detection system," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020, pp. 1723–1727.
[175] Y. Hu, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "DeFall: Environment-independent passive fall detection using Wi-Fi," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8515–8530, Jun. 2022.
[176] R. Hristov, Z. Kabelac, H. Rahul, and D. Katabi, "System and method for wireless detection and measurement of a subject rising from rest," U.S. Patent 17 235 447, Oct. 21, 2021.
[177] M. Gao, A. Souri, M. Zaker, W. Zhai, X. Guo, and Q. Li, "A comprehensive analysis for crowd counting methodologies and algorithms in Internet of Things," *Clust. Comput.*, vol. 27, pp. 859–873, Feb. 2024.
[178] S. Depatla and Y. Mostofi, "Crowd counting through walls using Wi-Fi," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, Athens, Greece, 2018, pp. 1–10.
[179] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Freecount: Device-free crowd counting with commodity Wi-Fi," in *Proc. GLOBECOM, IEEE Global Commun. Conf.*, Singapore, 2017, pp. 1–6.
[180] J. He and W. Yang, "IMep: Device-free multiplayer step counting with Wi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5887–5899, Oct. 2023.
[181] Y. Xu, W. Yang, J. Wang, X. Zhou, H. Li, and L. Huang, "WiStep: Device-free step counting with Wi-Fi signals," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–23, 2018.
[182] K. Qian, C. Wu, Z. Yang, C. Yang, and Y. Liu, "Decimeter level passive tracking with Wi-Fi," in *Proc. 3rd Workshop Hot Topics Wireless*, New York, NY, USA, 2016, pp. 44–48.
[183] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with Wi-Fi," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2329–2342, Nov. 2015.
[184] X. Liu, H. Chen, X. Jiang, J. Qian, G. Aceto, and A. Pescape, "Wi-CR: Human action counting and recognition with Wi-Fi signals," in *Proc. IEEE 4th Int. Conf. Comput., Commun. Security (ICCCS)*, Rome, Italy, 2019, pp. 1–8.
[185] D. Wu et al., "WiTraj: Robust indoor motion tracking with Wi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 3062–3078, May 2023.
[186] X. Guo, B. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "Wi-Fi-enabled smart human dynamics monitoring," in *Proc. 15th ACM Conf. Embed. Netw. Sensor Syst.*, Delft, The Netherlands, 2017, pp. 1–13.
[187] R. Arthi, D. S. Rawat, A. Pillai, Y. Nair, and S. Kausik, "Analysis of indoor localization algorithm for Wi-Fi using received signal strength," in *Proc. Int. Conf. Emerg. Trends Adv. Elect. Eng. Renew. Energy (ETAEERE)*, 2021, pp. 423–431.
[188] R. Estrada, N. Vera, and X. Aizaga, "Evaluation of machine learning prediction models for Wi-Fi based indoor positioning system," in *Proc. IEEE Int. Conf. Comput., Inf. Telecommun. Syst. (CITS)*, Genoa, Italy, 2023, pp. 01–07.
[189] M. Mallik, S. Das, and C. Chowdhury, "Rank based iterative clustering (RBIC) for indoor localization," *Eng. Appl. Artif. Intell.*, vol. 121, May 2023, Art. no. 106061.
[190] G. Caso, L. De Nardis, F. Lemic, V. Handziski, A. Wolisz, and M.-G. Di Benedetto, "ViFi: Virtual fingerprinting Wi-Fi-based indoor positioning via multi-wall multi-floor propagation model," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1478–1491, Jun. 2020.
[191] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.
[192] V. Narayan, S. Awasthi, N. Fatima, M. Faiz, and S. Srivastava, "Deep learning approaches for human gait recognition: A review," in *Proc. IEEE Int. Conf. Artif. Intell. Smart Commun. (AISC)*, 2023, pp. 763–768.
[193] C. Wu, F. Zhang, Y. Hu, and K. R. Liu, "GaitWay: Monitoring and recognizing gait speed through the walls," *IEEE Trans. Mobile Comput.*, vol. 20, no. 6, pp. 2186–2199, Jun. 2021.

[194] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using Wi-Fi signals," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, Heidelberg, Germany, 2016, pp. 363–373.

[195] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: Wi-Fi-based person identification in smart spaces," in *Proc. 15th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Vienna, Austria, 2016, pp. 1–12.

[196] Y. Zhou, J. Yang, H. Huang, and L. Xie, "AdaPose: Towards cross-site device-free human pose estimation with commodity Wi-Fi," 2023, *arXiv:2309.16964*.

[197] J. Yang, Y. Zhou, H. Huang, H. Zou, and L. Xie, "MetaFi: Device-free pose estimation via commodity Wi-Fi for metaverse avatar simulation," in *Proc. IEEE 8th World Forum Internet Things (WF-IoT)*, Yokohama, Japan, 2022, pp. 1–6.

[198] Y. Zhang, X. Wang, J. Wen, and X. Zhu, "Wi-Fi-based non-contact human presence detection technology," *Sci. Rep.*, vol. 14, no. 1, p. 3605, 2024.

[199] C. Li, Z. Cao, and Y. Liu, "Deep AI enabled ubiquitous wireless sensing: A survey," *ACM Comput. Surveys*, vol. 54, no. 2, pp. 1–35, 2021.

[200] Q. Li, R. Gravina, Y. Li, S. H. Alsamhi, F. Sun, and G. Fortino, "Multi-user activity recognition: Challenges and opportunities," *Inf. Fusion*, vol. 63, pp. 121–135, Nov. 2020.

[201] F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, "A survey on deep learning for human activity recognition," *ACM Comput. Surveys*, vol. 54, no. 8, pp. 1–34, 2021.

[202] T. Koike-Akino, P. Wang, M. Pajovic, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial mmWave Wi-Fi: A deep learning approach," *IEEE Access*, vol. 8, pp. 84879–84892, 2020.

[203] R. F. Miranda et al., "A review of cognitive hybrid radio frequency/visible light communication systems for wireless sensor networks," *Sensors*, vol. 23, no. 18, p. 7815, 2023.

[204] C. Liang et al., "Integrated sensing, lighting and communication based on visible light communication: A review," *Digit. Signal Process.*, vol. 145, Feb. 2024, Art. no. 104340.

[205] T. Li, C. An, Z. Tian, A. T. Campbell, and X. Zhou, "Human sensing using visible light communication," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Paris, France, 2015, pp. 331–344.

[206] K. Gligorić, M. Ajmani, D. Vukobratović, and S. Sinanović, "Visible light communications-based indoor positioning via compressed sensing," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1410–1413, Jul. 2018.

[207] H. Yang, W.-D. Zhong, C. Chen, and A. Alphones, "Integration of visible light communication and positioning within 5G networks for Internet of Things," *IEEE Netw.*, vol. 34, no. 5, pp. 134–140, Oct. 2020.

[208] A. Ullah, W. Choi, and S. Coleri, "Path loss estimation and jamming detection in hybrid RF-VLC vehicular networks: A machine-learning framework," *IEEE Sensors J.*, vol. 23, no. 24, pp. 31325–31336, Dec. 2023.

[209] X. Yuan, Y.-J. A. Zhang, Y. Shi, W. Yan, and H. Liu, "Reconfigurable-intelligent-surface empowered wireless communications: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 136–143, Apr. 2021.

[210] A. Kaushik et al., "Towards integrated sensing and communications for 6G: A standardization perspective," 2023, *arXiv:2308.01227*.

[211] Y. Han, W. Tang, X. Li, M. Matthaiou, and S. Jin, "CSI acquisition in RIS-assisted mobile communication systems," *Nat. Sci. Rev.*, vol. 10, no. 8, 2023, Art. no. nwad127.

[212] J. Hu et al., "Reconfigurable intelligent surface based RF sensing: Design, optimization, and implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2700–2716, Nov. 2020.

[213] A. Aubry, A. De Maio, and M. Rosamilia, "RIS-aided RADAR sensing in N-LOS environment," in *Proc. IEEE 8th Int. Workshop Metrol. AeroSpace (MetroAeroSpace)*, Naples, Italy, 2021, pp. 277–282.

[214] L. Li, K. Ota, and M. Dong, "Human in the loop: Distributed deep model for mobile crowdsensing," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4957–4964, Dec. 2018.

[215] L. Duan, L. Huang, C. Langbort, A. Pozdnukhov, J. Walrand, and L. Zhang, "Human-in-the-loop mobile networks: A survey of recent advancements," *J. Sel. Areas Commun.*, vol. 35, no. 4, pp. 813–831, 2017.

[216] A. T. Z. Kasgari, W. Saad, and M. Debbah, "Human-in-the-loop wireless communications: Machine learning and brain-aware resource management," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7727–7743, Nov. 2019.

IFTIKHAR AHMAD (Student Member, IEEE) received the bachelor's and master's degrees in telecommunication engineering from the University of Engineering and Technology (UET) Peshawar and UET Taxila, Pakistan, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Smart Networking Laboratory, Department of Computer Engineering, College of IT Convergence, Chosun University, Republic of South Korea. His research interest includes machine-learning-assisted RF-based HAR.

ARIF ULLAH (Member, IEEE) received the M.S. degree in electrical engineering from COMSATS University, Islamabad, Pakistan, and the Ph.D. degree in electronic and communication engineering from Ghulam Ishaq Khan (GIK) Institute of Engineering Sciences and Technology, Topi, Pakistan. He was a Graduate Research Assistant with GIK Institute and also worked as a Research Fellow with the Smart Networking Laboratory, Chosun University, South Korea, from November 2021 to April 2022. He is currently working as an Assistant Professor with the Department of Computer Engineering, College of IT Convergence, Chosun University. His research interests include modeling and performance analysis of cellular networks using stochastic geometry, millimeter wave, and terahertz communication, vehicular networks, Internet of Things, 5G and beyond 5G networks, and machine-learning-aided wireless communication.

WOOYEOL CHOI (Member, IEEE) received the B.S. degree from the Department of Computer Science and Engineering, Pusan National University, Busan, Republic of Korea, in 2008, and the M.S. and Ph.D. degrees from the School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea, in 2010 and 2015, respectively. He is currently an Associate Professor with the Department of Computer Engineering, Chosun University, Gwangju. He was a Senior Research Scientist with the Korea Institute of Ocean Science and Technology, Ansan, Republic of Korea, from 2015 to 2017, and a Senior Researcher with Korea Aerospace Research Institute, Daejeon, Republic of Korea, from 2017 to 2018. His research interests include cross-layer protocol design, learning-based resource optimization, and experiment-driven evaluation of wireless networks.