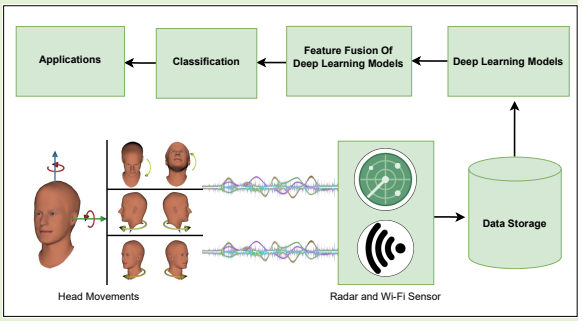
Wi-Fi and Radar Fusion for Head Movement Sensing Through Walls Leveraging Deep Learning

Hira Hameed, *Student Member, IEEE*, Ahsen Tahir, *Member, IEEE*, Muhammad Usman, *Member, IEEE*, Jiang Zhu, *IEEE, Fellow*, Lubna, *Student Member, IEEE*, Hasan Abbas, *Member, IEEE*, Naeem Ramzan, *Senior Member, IEEE* , Tie Jun Cui, *IEEE, Fellow*, Muhammad Ali Imran, *IEEE, Fellow*, and Qammer H. Abbasi, *Senior Member, IEEE*

***Abstract*— The detection of head movement plays a crucial role in human-computer interaction systems. These systems depend on control signals to operate a range of assistive and augmented technologies, including wheelchairs for Quadriplegics, as well as virtual/augmented reality and assistive driving. Driver drowsiness detection and alert systems aided by head movement detection can prevent major accidents and save lives. Wearable devices, such as MagTrack consist of magnetic tags and magnetic eyeglasses clips and are intrusive. Vision-based systems suffer from ambient lighting, line of sight, and privacy issues. Contactless sensing has become an essential part of next-generation sensing and detection technologies. Wi-Fi and radar provide contactless sensing, however in assistive driving they need to be inside enclosures or dashboards, which for all practical purposes in this paper have been considered as through walls. In this study, we propose a contactless system to detect human head movement with and without walls. We used ultra-wideband(UWB) radar and Wi-Fi signals, leveraging machine and deep learning techniques. Our study analyzes the six common head gestures: right, left, up, and down movements. Time-frequency multi-resolution analysis based on wavelet scalograms is used to obtain features from channel state information values, along with spectrograms from radar signals for head movement detection. Feature fusion of both radar and Wi-Fi signals is performed with state-of-the-art deep learning models. A high classification accuracy of 83.33% and 91.8% is achieved overall with the fusion of VGG16 and InceptionV3 model features trained on radar and Wi-Fi time-frequency maps with and without the walls, respectively.**



***Index Terms*— RF sensing, behavior monitoring, micro-Doppler signatures, channel state information, deep learning, machine learning, features fusion.**

I. Introduction

HEAD movements [1], carry important information related to human behavior. Head motions are an integral part of non-verbal communication and have a wide range of applications for human-computer interaction, such as assistive technologies, virtual and augmented reality, and assistive driving systems. Head movement detection has been widely utilised for assistive driving of wheelchairs for patients suffering from paralysis, driver drowsiness detection and alert systems. Intelligent assistive driving systems can reduce the number of road accidents by monitoring driver's behavior through head movements and generate alerts accordingly. Mental tiredness impairs focus when driving and has major safety implications [2], [3]. Poor sleep and tiredness are major causes of poor driving performance, steering mistakes, loss of vehicle control, and deadly accidents [4]–[7]. Driving assistance systems rely heavily on the detection of driver attentiveness. The orientation of the driver's head may reflect his degree of attention. Head movement is getting high popularity in assistive driving since an estimated 1,560 reported road deaths in 2021 in the UK [8]. In recent years, there has been an increase in assistive technologies in healthcare and many other domains that benefit from smart technology concepts. Head movement detection has proven to be effective in many applications such as the detection of driver's fatigue [9], human visual focus [10], behavior recognition [11], vitals monitoring [12], healthcare cognitive assistance [13], in figuring out the human head kinematics [14] to estimate and predict possible

Hira Hameed, Hasan Abbas, Muhammad Ali Imran, and Qammer H. Abbasi are with University of Glasgow, James Watt School of Engineering, Glasgow, G12 8QQ, UK ({Hira.Hameed, hasan.abbas, muhammad.imran, qammer.abbasi}@glasgow.ac.uk).

Muhammad Usman is also with the School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow, G4 0BA, UK (muhammad.usman@gcu.ac.uk).

Ahsen Tahir is also with Department of Electrical Engineering, University of Engineering and Technology, Lahore, PK (ahsan@uet.edu.pk).

Lubna is with Department of Telecommunication Engineering Dept. UET Peshawar (Lubnaxafi@gmail.com).

Naeem Rizwan is with the School of Computing, Engineering and Physical Sciences, University of West of Scotland, UK (Naeem.Ramzan@uws.ac.uk).

Tie Jun is with School of Information Science and Engineering. Southeast University, Nanjing, China (tjcui@seu.edu.cn).

Jiang Zhu is with Meta, USA (jiangzhu@ieee.org).

Qammer H. Abbasi is with the Artificial Intelligence Research Center (AIRC), Ajman University, UAE

head collision injuries in athletes, and in clinical depression monitoring [15] etc. Real-time head movements estimation techniques are also being integrated with mobile devices which can assist in multiple healthcare applications. Thread-based sensors are also used along with machine learning algorithms to classify various head movements [16]. This research [16] describes a method for tracking and classifying head movements using flexible strain-sensing threads attached to the neck of an individual. A data processing technique for motion recognition quantifies head location in near real-time. To predict the location, a set of features is extracted from each data segment and utilized as input for nine classifiers, including Support Vector Machine, Naive Bayes, and KNN. Several other techniques to estimate head position are surveyed in [17]. Using multi-primitive closed-loop face analysis in video arrays, [18] developed a computational framework for robust face identification and posture estimation. [19] used facial symmetry and anthropometric measurements to compute head orientation. The head's Y-Z coordinates were calculated using eye distances and camera focal length. Face anthropometry was used to estimate head X-axis orientation. This method was tested on actual photos. [20] presented a real-time head movement estimation approach based on the video camera as a way of communication between the individual and the device. The suggested solution is made up of numerous computer-vision algorithms that have been carefully tuned to work in a specific environment, as well as a head posture estimation based on rolling/yaw, and pitching movement calculations. Experiments were carried out using 363 videos of 27 individuals in various settings. Also, camera-based and wearable devices were used to recognize head movements, which were discussed in the literature for identifying human behavior while listening, talking, and in driving assistance applications. These techniques have limitations, such as the obligation to record the target, which restricts their practical uses due to privacy concerns. The legal implications of such aids may restrict their wider use in public and private settings; for example, video-in-head motions may be viewed as photographing someone without their consent, which is illegal in many countries. The main drawbacks of existing camera-based and wearable-based technology include serious privacy concerns, poor lighting, obstructions to the line of sight, training difficulties with longer video sequence data, and computational complexities, and wearable devices disrupt daily routines.

Radio frequency (RF) head movement sensors, on the other hand, can fulfill the demand for next-generation technologies. By recognizing head motions using RF sensing, machine learning (ML), and deep learning (DL) techniques, various applications can benefit from very accurate cues. Moreover, unlike vision-based systems, RF sensing-based head movements are unaffected by opaque barriers or walls separating the target and the transponder. RF signals can pass through walls to detect visual cues, such as head and lip movements. Head movements provide additional functionality for the next generation of multimodal hearing aid devices for understanding the behavior of people. In this study, we designed, developed, and tested an RF sensing-based method for detecting head motions with and without a wall. Activity monitoring through walls or barriers via Wi-Fi and radar devices is a great breakthrough in the field. Since cameras are limited to line-of-sight visuals and they can not detect/sense any object or humans through walls/barriers. Therefore in this work, we introduced a radar and Wi-Fi-based novel system which can perform head-movements monitoring through walls and other opaque barriers.

These papers [21]–[23], provide an explanation of all applications conducted through RF (Radio Frequency) covering speech recognition, activity recognition, and hand gestures with high accuracies. The advantages and challenges of radar-based driving assistance systems are presented in [24]. For the automotive-radar industries, the main system requirements are to achieve high-resolution, low-cost hardware and size compactness. The major strength of using automotive-Radars in driving-assistance systems is the higher angular resolutions attained even with a small number of antennas being used. The authors also discussed the high-resolution angle-finding techniques that are computationally effective for automotive-radar applications. In another work, [25], four human activity recognition such as (a) box (punch forward three times), (b) pick (squat down and pick something up), (c) foot (four steps in place), and (d) zombie (raise a hand like a zombie) were performed through the wall using radar and achieves 97.6% accuracy. Similarly, the authors in [26] presented human activity (walking, sitting, and falling) detection system using Wi-Fi signals. In this work, the transmitter and receiver were separated by the wall and activities were performed on both sides of the transmitter and receiver side.

Our work focuses on recognizing different head movements and collecting data using micro-Doppler signatures and CSI amplitude using a radar sensor and Wi-Fi signals. The existing dataset is diverse in nature that includes samples from a wide variety of subjects (ages and genders) and a diverse number of classes that cover all essential aspects of head movements. Head up, Head down, Right 90, Left 90, Right 45, and Left 45 are the six types of Doppler signatures and CSI data considered for this work. These types of movements include dynamic gestures in which mobility or head are used to represent various movements. The dataset was recorded using two separate methods i.e. using a Radar sensor and Wi-Fi signals with and without a wall. These features make the dataset a better option for the training and assessment of ML and DL algorithms for the recognition of head movements. In order to visualize the recorded data, spectrograms and CSI amplitudes were used.

The following presents the main contributions of our research work in the field:

- We proposed a contact-less head recognition system that automatically recognizes and translates head movements with and without a wall in between the target and transponder setup.
- In addition, we collect a dataset of 2400 samples from 6 different types of head movements captured at 0.50 centimeters distance away from the target. Furthermore, the data samples are collected using 2 different techniques (Radar sensor and Wi-Fi signals) with and without a wall. To ensure diversity, data was collected from four participants (two males and two females) ranging in age

from 20 to 40 years.

- For the radar dataset, VGG16, VGG19, InceptionV3, and SqueezeNet were applied on the individual subject, combined dataset of four subjects VGG16 outperformed as compared to another algorithm with 80% Accuracy with the wall and 79.2% without the wall.
- For the Wi-Fi dataset, VGG16, VGG19, InceptionV3, SqueezeNet, Neural network pattern recognition, Tree(Medium Tree), and Ensemble(Boosted Tree) was applied on the individual subject, a combined dataset of four subjects InceptionV3 outperformed as compared to another algorithm 80% Accuracy with the wall and 89% without the wall.
- The fusion of features for different deep learning models was tested. The highest accuracy values of 91% without the wall were achieved with feature fusion of VGG16 and InceptionV3 deep learning models. Furthermore, the highest accuracy of 83.33% was achieved through the walls with the feature fusion of VGG16 and InceptionV3 deep learning models.
- In this work, we presented the experimental results from several state-of-the-art DL and ML models applied to our benchmark dataset, which can serve as a foundation for future research in the domain of detecting head movements through walls.

This research proposes novel head movement gestures using micro-doppler signatures using radar-sensor with and without walls. Five different gestures are considered Head 45L, Head 45R, Head 90L, Head 90R, and Head Down. An ultra-wideband radar, XeThru X4M03 is used to record experimental data. The received data is represented in the form of spectrograms while spatiotemporal features were extracted using fusion of two different models. We achieved 91.8% of classification accuracy without a wall. The possible use cases of the proposed technology are illustrated in Fig. 1. The whole setup, data collection, DL and ML algorithms, and experimental results are presented in the following sections.

II. Radar based setup

The experimental setup and configuration parameters for the radar-based head movement system are illustrated in Fig. 2a. The sensor has a 1.5 GHz sensor bandwidth and a detection range of about 9.6 meters. It utilizes an ultra-wide-wide (UWB) radar sensor, specifically the Xethru-X4M03 model, equipped with both transmitter (Tx) and receiver (Rx) antennas. To derive valuable insights from the radar data, we employed the short-time Fourier transform (STFT) on the radar signal. This process resulted in the creation of spectrograms that effectively captured the radar Doppler shift corresponding to various head movements. Examination of these spectrograms revealed that different head movements produced distinct spectrogram patterns.

A. Scenario 1 - Line-of-sight: With no wall in between target and transponder setup

The sensor was placed in front of the participants/subject at around a half-meter distance. The experimental data recording activity for head movements was carried out by placing the radar 0.5 meters away from the subject sitting on a chair. The only movements performed here by the subjects were the head movements with slight shoulder movements which naturally arises while talking. The rest of the body was in a normal sitting position. Each activity was performed in a 4 seconds time frame. In these 4 seconds, the RF signal was transmitted and received back by the radar. The data collection and processing using UWB radar setup are shown in Fig. 3a.

B. Scenario 2 -Non-line-of-sight: With wall/opaque barrier in between target and transponder setup

The sensor was placed in the line of sight of the participants/subject at around a half-meter distance. A plasterboard/drywall wall was placed between the target and the radar. The experimental data recording activity for each head movement was carried out for 4 seconds and during these four seconds, the radar sent and received the RF signals. The subject was sitting on the chair in a normal position while performing head movements activity. The data collection and processing using UWB radar setup are shown in Fig. 3b.

III. Wi-Fi based setup

The second set of experiments was performed using Wi-Fi. The experimental setup and parameter configuration for Wi-Fi based head movement system is given in Fig. 2b. The main equipment of this setup is USRP-X300 with a single transmitter antenna (directional) and two antennas at the receiver side which are (directional) in nature. On the transmitter side, the Rx antenna UWB 1.35GHz-9.5GHz Log-Periodic Directional was used as a transmitter whereas two monopole antennas (VERT2450) optimized at an operating frequency of 5.5 GHz were used as a receiver. The gain of both the Tx/Rx antennas was set to 35 dB. The USRP and desktop were connected using an Intel(R)-Core(TM) i7-7700 processor operating at 3.60 GHz with 16GB of RAM. Communication between the USRP and GNU Radio was established using a virtual machine running Ubuntu 16.04. A Python script was employed to transmit and receive data from the USRP-X300. The experiments were conducted within the 5.5 GHz Wi-Fi frequency band.

A. Scenario 1 - Line-of-sight: With no wall in between target and transponder setup

The Tx and Rx antennas were situated approximately 0.50 meters away from the subject, and each head movement was performed continuously for 4 seconds. The data collection process and subsequent processing using Wi-Fi equipment are depicted in Fig. 3c. It's important to note that Wi-Fi signals were evaluated based on a range of characteristics, including time-frequency maps, among others. Unlike radar signals, where frequency shift was the primary distinguishing factor, Wi-Fi CSI values were most effective when variations in CSI amplitudes were observed. These fluctuations in one-dimensional CSI amplitude revealed distinctive patterns of head movement.

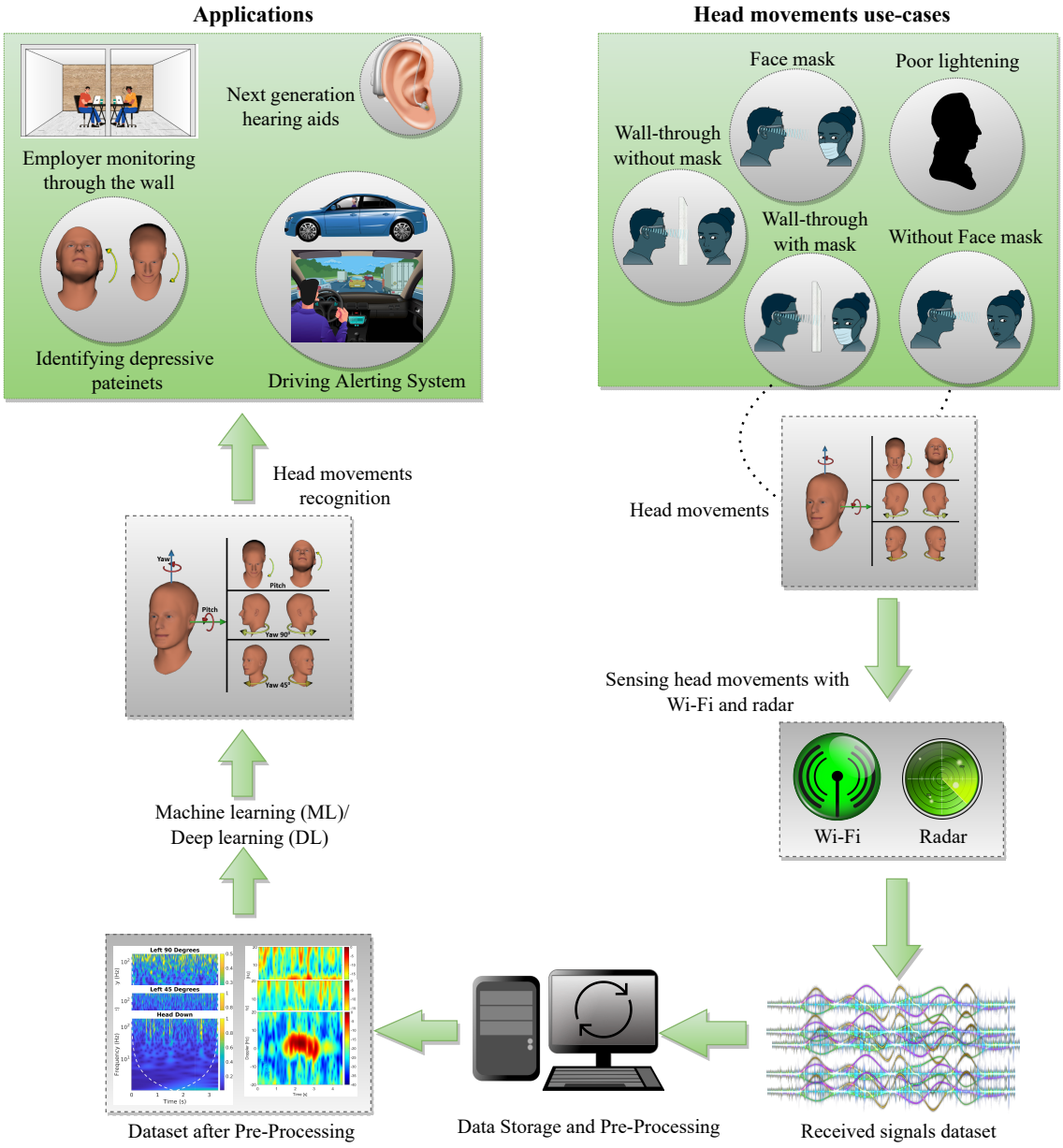


Fig. 1. Conceptual representation of the suggested methodology for head movements.

B. Scenario 2 - Non-line-of-sight: With wall/opaque barrier in between target and transponder setup

The Tx and Rx antennas were positioned around 0.50 meters away from the subject. Plasterboard or drywall was placed between Tx, Rx, and target. The experimental data recording activity for each head movement was carried out for 4 seconds and during these four seconds, the Tx signal hit the target and was received back to the receiver. The subject was sitting on the chair in a normal position while performing head movements activity. The data collection and processing using Wi-Fi setup are shown in Fig. 3d.

IV. Methods

The main illustration of head movement activity is shown in Fig. 4a. In the case of Wi-Fi, 2000 packets were transmitted within four seconds, where each data instance represented the CSI amplitudes. The CSI patterns (amplitude) of considered head movements, namely, Head down, Head up, Head left 90, Head Right 90, Head Right 45, and Head Left 45, are depicted in Fig. 4b without wall and Fig. 4d with wall experiments. In each figure, the 51 subcarriers of the OFDM signal are represented by different colors. The amplitude of the subcarriers is represented on the y-axis of each sub-figure, while the number of received packets is displayed on the x-axis. In the radar scenario, The same approach was used for data collection with a total of 600 data samples, four subjects participated including two males and two females, with 25 data samples in each class. Data is in the form of a spectrogram, which is shown in Fig. 4c without wall and Fig. 4e with wall experiments. Each figure's different colors represent a

a

Parameter	Value
Platform	Xetru radar X4MO3
Instrumental range	9.6 meters
Subject and Radar distance	0.50 meters
Frequency of operation	7.29GHz
Tx power	6.3dBm
Activity duration	4 seconds
Collected samples in each class	25

b

Parameter	Value
USRP-Platform	X300
OFDM-subcarriers	51
Frequency of operation	5.5GHz
Gain of Tx	35dB
Gain of Rx	35dB
Tx antenna	Log periodic HyperLOG 7040, 700MHz to 4GHz
Rx antenna	UWB 1.35GHz-9.5GHz Log-Periodic Directional
Subject distance from Tx-Rx antennas	0.50 meters
Duration of activity	4 seconds
Samples collected (each class)	50

c

Subject		Head Movements												
		Radar						Wi-Fi						
		Head down	Head Up	Left 45	Left 90	Right 45	Right 90	Head down	Head Up	Left 45	Left 90	Right 45	Right 90	Total
S1 (Male)	With-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
	Without-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
S2 (Male)	With-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
	Without-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
S3 (Female)	With-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
	Without-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
S4 (Female)	With-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
	Without-Wall	25	25	25	25	25	25	25	25	25	25	25	25	300
Total		200	200	200	200	200	200	200	200	200	200	200	200	2400

Fig. 2. Head movements activity with their representation in Wi-Fi and radar signal. (a) The configuration parameters of radar software and hardware without and with through the wall experiment. (b) The configuration parameters of Wi-Fi software and hardware with and without the wall experiment. (c) An overview of the gathered data, the total number of participants, and the conducted activities.

Fig. 3. Head movements activity with their representation in Wi-Fi and radar signal. (a) An experimental setup of the radar signal without a wall. (b) An experimental setup of radar signal through a wall. (c) An experimental setup of Wi-Fi signal without wall. (d) An experimental setup of Wi-Fi signal through a wall.

change in frequency. In each spectrogram, y-axis represents the Doppler shift (Hz), while the x-axis represents time.

A. Radar Data Processing

The Xethru X4M03 radar chip was configured using the XEP interface and X4driver. Data was recorded at a rate of 500 frames per second (FPS) in the form of float message data. A loop was implemented to read the data file, and the values were subsequently stored in a DataStream variable, which was then converted into a complex range-time-intensity matrix. To generate a Doppler range map, a moving target indication (MTI) filter was applied. Spectrograms were created using the following parameters: overlap percentage, window length set to 128, fast Fourier transform (FFT), and padding factor set to 16. A second MTI filter, a Butterworth 4th-order filter, was used. Each chirp underwent an initial FFT transformation to produce a range profile. Subsequently, a second FFT was performed on a specific number of chirps in a sequence for each range bin. Spectrograms were created using the Short-Time Fourier Transform (STFT), which segments the data and applies the Fourier transform to each segment, providing information about both time and frequency. The radar data's

Doppler information depends on the hardware sampling rate, and the highest unambiguous Doppler frequency in radar is determined by the chirp time, given by the formula $F_d, max = \frac{1}{2t_c}$.

Head movements recognition at a distance D(t) from a specified location such as the head is the focus of this paper.Ts is the transmitted signal, while $V(t)$ is the target position in front of the RADAR,

$$T_s(t) = E\cos(2\pi ft). \tag{1}$$

The signal received is provided by Rs(t),

$$R_s(t) = \acute{E}\cos(2\pi f(t - \frac{2D(t)}{c})), \tag{2}$$

where the speed of light is c and E is the reflection coefficient. The signal that is reflected off the target points at an angle $theta$ to the direction of the RADAR and is denoted by the symbol $Rs(t)$.

$$R_s(t) = \acute{E}\cos(2\pi f(1 + \frac{2v(t)}{c})(t - \frac{4\pi D(\theta)}{c})). \tag{3}$$

The corresponding Doppler shift can be expressed as,

$$f_d = f\frac{2v(t)}{c}. \tag{4}$$

The signal that is received back is composed of a number of moving parts, including the head and other small motions of the body. Each component moves with its own acceleration and speed. The received signal can be written as if i were the various moving parts of the head. We can write as

$$R_s(t) = \sum_{k}^{N} A_k\cos(2\pi f(1 + \frac{2vk(t)}{c})(t - \frac{4\pi D_k(0)}{c})). \tag{5}$$

The Doppler shift is the result of a complex interaction of multiple Doppler shifts due to various head movements. The feature of doppler signatures depends on the detection of head movements. After getting the spectrogram of different subjects, It was divided into two datasets: (i) Data Training and (ii) Data Testing. The spectrogram fed into the proposed pre-trained DL classification algorithm for the classification of the head movements dataset.

B. Wi-Fi Data Processing

The data was transmitted using OFDM symbols with 52 subcarriers that were tightly spaced. As shown by Eq. 6, data were collected in a matrix form having frequency responses of subcarriers N=51.

$$H = [H_1(f), H_2(f), \cdots, H_M(f)]^K, \tag{6}$$

Here, the Hl-frequency subcarrier can be expressed as

$$H_l\,(f) = |H_l\,(f)\,|e^{l\angle H_l(f)}, \tag{7}$$

where, amplitude $|H_l(f)|$ and phase $\angle H_l(f)$ are responses of the lth subcarrier. All subcarrier responses correlated with system input and output as shown in Eq. 8,

$$H_k(f) = \frac{Y_l(f)}{X_l(f)}, \tag{8}$$

where input and output Fourier transformations are denoted by $Xl(f)$ and $Yl(f)$, respectively. The received CSI data often contain environmental noise. Therefore, the collected data is processed by eliminating the mean received power for each subcarrier from every sample. To observe the maximum variation due to head movements, the subcarrier with the highest variance was identified for feature extraction. These 10 features were extracted from the dataset namely minimum, median, variance, eight peaks, standard deviation, high order moments, mode, skewness, kurtosis, and moments. After taking features which were in comma-separated form (CSV) file, which was used to train various machine learning algorithms that are described in other section. After that, to accurately classify the head movement classes, training, and testing were carried out using the test-train split evaluation method.

V. Parameter Settings of the ML and DL Algorithms

The presented head movements classification approach was divided into two parts: (i) system training and (ii) system testing: For Wi-Fi dataset ML algorithms such as NN(Neural network pattern) recognition, Tree(Medium tree), and Ensemble(Boosted trees) were applied. while on radar data, the VGG16, VGG19, InceptionV3, and SqueezeNet. DL pre-trained models were applied to the radar data-generated spectrogram images.

The ML and DL model parameter settings are shown in the table. I.

VGG16 Model: The data was input into the VGG16 models convolution layers with rectified linear unit (ReLU) activation functions and 3×3 kernel sizes. Each convolution layer was followed by a max-pooling layer with 2×2 kernel sizes. The final layer comprised three fully connected layers (FC). The convolution layer and FC layers contained the training weights, which determined the number of parameters.

VGG19 Model: The data was passed through a different layer which consists of 3×3 filters with five stages of convolutional layers, five pooling layers, and three fully connected layers to get image information. The convolution kernel depth has been increased from 64 to 512 of the VGG16 network for better image feature vector extraction. Every stage of convolutional layers was followed by pooling layers which have the size and step size of 2×2.

InceptionV3 Model: The dataset was processed using the InceptionV3 deep learning model, which consists of 48 layers. The architecture of the model involves a sequence of three convolution layers, followed by a max pooling layer, two more convolution layers, and another max pooling layer. Spectrograms were input into the model, which then underwent multiple convolutions using various filters. This process was repeated several times across the entire network to facilitate image classification.

SqueezeNet Model: SqueezeNet is an 18-layer deep convolutional neural network. Spectrograms of the input were sent to

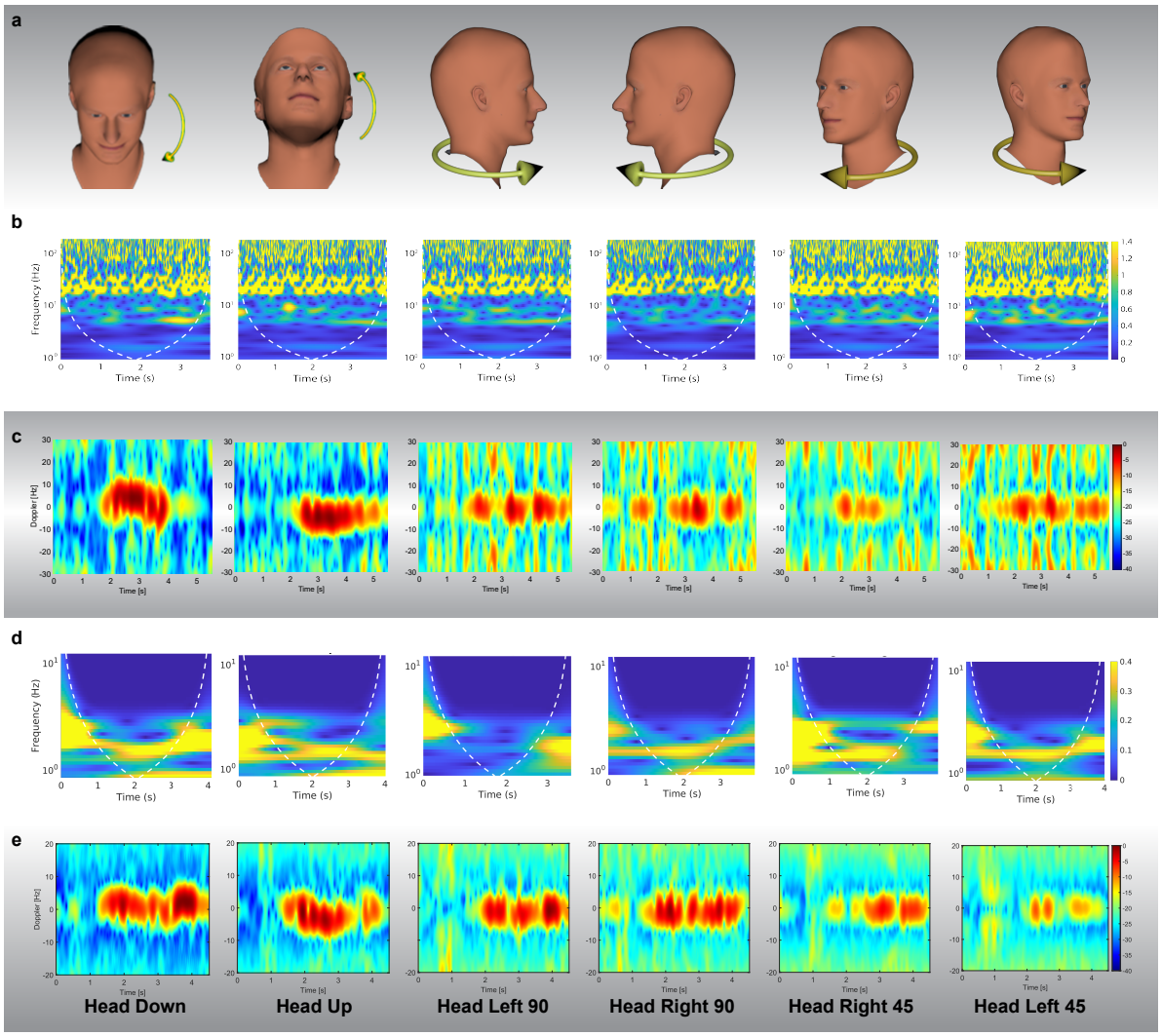


Fig. 4. Wi-Fi and radar signal representation of head movement activity. (a) A visual representation of head movements from various angles. (b) Wi-Fi data samples representing various classes of head movements without walls. (c) Radar data samples representing various Head movement classes without the wall. (d) Wi-Fi data samples representing various classes of head movements with walls. (e) Data samples from radar that represent different head movement classes with the presence of a wall.

the layers. The last convolution layers were added as follows the dropout layer was set to 50%, convolution layers with stride, Relu as activation function, Global average pooling, and softmax layer were added before the classification output layer.

NN (Neural Network Pattern Recognition) Model: The pattern recognition neural network used in this study comprises two-layer feed-forward networks with hidden neurons using sigmoid activation functions. SoftMax activation functions were applied to the output layer neurons. The network was trained using the scaled conjugate gradient backpropagation algorithm, which involved updating the weight and bias values as data passed through these layers. Subsequently, the dataset was partitioned into training, validation, and testing subsets. The network's performance was assessed based on cross-entropy and misclassification error metrics.

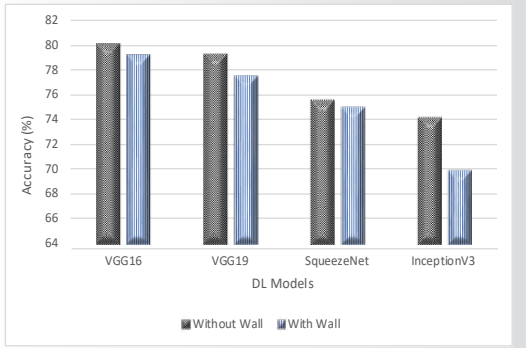
Tree (Medium Tree) Model: Data were fed to decision trees, classification trees, and regression trees for classification. It followed the decisions in the trees down to a leaf node in order to forecast a reaction. The response was located in the leaf node. Classification trees provided nominal answers, such as "true" or "false".

Ensemble (Boosted Tree) Model: The classifier has the ability to combine the results of multiple low-quality learners into a single high-quality model. The data were input to the booting ensemble algorithm, which identified the highest breakpoints or branch points to handle the depth of tree learners. The experimental setup achieved improved precision with a learning rate of 0.1.

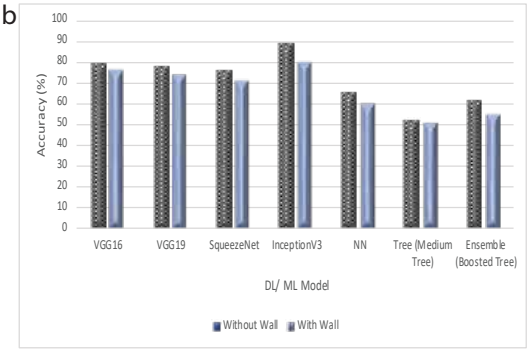
VI. Results and Discussion

Two RF sensing technologies were used in two different experiments with and without a wall, *i.e.,* Wi-Fi and radar. Data collection involved the capture of six head movements:

a



b



c

DL Model	TPR/FPR (%)	Data Fusion						
		Head down	Head Up	Left 45	Left 90	Right 45	Right 90	Accuracy (%)
VGG16 + InceptionV3	TPR (%)	100	100	86.0	81.0	80.0	87.0	91.8
	FPR (%)	0.0	0.0	14.0	29.0	20.0	13.0	

d

DL Model	TPR/FPR (%)	Data Fusion						
		Head down	Head Up	Left 45	Left 90	Right 45	Right 90	Accuracy (%)
VGG16 + InceptionV3	TPR (%)	96.0	95.0	89.0	86.0	76.0	58.0	83.33
	FPR (%)	4.0	5.0	11.0	14.0	24.0	42.0	

Fig. 5. Overall system overview and the results. (a) The comparative result of radar-based system with and without wall using Deep Learning models. (b) The comparative result of a Wi-Fi-based system with and without wall using Deep and Machine Learning models. (c) The data fusion result of Wi-Fi and radar data without wall using deep learning models. (d) The data fusion result of Wi-Fi and radar data through the wall using deep learning models.

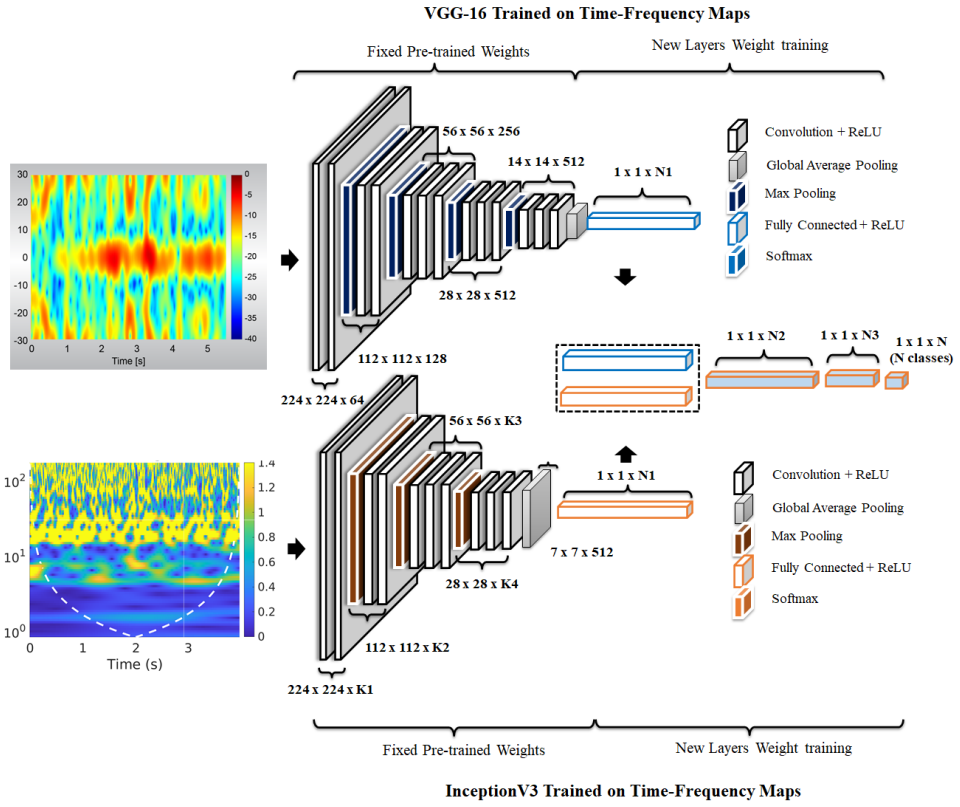


Fig. 6. Feature fusion of radar and Wi-Fi time-frequency maps.

DL/ML Model	Parameters	Settings
VGG16	Number of Layers	16
	Learning rate	0.0001
	Batch size	16
	Learning algorithm	Adam
	Loss function	Cross entropy
	Number of epochs	25
VGG19	Number of Layers	19
	Learning rate	0.0001
	Batch size	16
	Learning algorithm	Adam
	Loss function	Cross entropy
	Number of epochs	25
InceptionV3	Number of Layers	48
	Learning rate	0.0001
	Batch size	16
	Learning algorithm	Adam
	Loss function	Cross entropy
	Number epochs	25
SqueezeNet	Number of Layers	18
	Learning rate	0.0001
	Batch size	16
	Learning algorithm	Adam
	Loss function	Cross entropy
	Number of epochs	25
NN	Number of Layers	10
	Training Function	Scaled conjugate
	Number of epochs	Gradient Backpropagation
	Loss function	20
		Cross entropy
Tree (Medium Tree)	SplitCriterion	gdi
	MaxNumSplits	20
	Surrogate	off
	KFold	5
	Loss Function	Classiferror
Ensemble	Learner type	Decision Tree
	Ensemble Method	AdaBoost
	Loss Function	Classiferror
	Learning rate	0.1
	Number of learners	30
	Maximum Number of splits	20

TABLE I
PARAMETER SETTINGS FOR THE SELECTED DEEP AND MACHINE LEARNING MODELS

Head up, Head down, Head Right 90, Head Left 90, Head Right 45, and Head Left 45. These movements were recorded with subjects in a stationary position and their bodies in a typical posture. To enhance the dataset's authenticity, four participants (two males and two females) took part in both the radar and Wi-Fi experiments. A total of 2400 data samples were collected from both experiments using radar and Wi-Fi, with and without a wall for six classes namely Head up, Head down, Head Right 90, Head Left 90, Head Right 45, and Head Left 45 which is shown in Fig. 2c. In each experiment with wall and without wall using radar, a total of 600 data samples were collected from four participants, where 25 samples were taken from each class. Specifically, each participant repeated each head movement activity 25 times with the radar. Likewise, the same number of data was acquired from USRP using the same strategy. The University of Glasgow's Research Ethics Committee granted ethical approval for these experiments (approval no.: 300200232, 300190109). In the context of radar datasets, both with and without a wall, the evaluation outcomes for the considered DL algorithms (VGG16, VGG19, SqueezeNet, and InceptionV3) are presented in Figure 5a. Notably, all the algorithms produced comparable results, with VGG16 slightly outperforming the others in both scenarios, whether with or without a wall, when using a combined dataset. Specifically, when employing the VGG16 algorithm, a classification accuracy of 80.0% is achieved on the combined dataset without a wall, which is marginally reduced to a promising accuracy of 79.2% in the presence of a wall. The evaluation results for various DL and ML algorithms (including VGG16, VGG19, SqueezeNet, InceptionV3, Neural network pattern recognition, Tree (Medium Tree), and Ensemble (Boosted Tree)) for Wi-Fi signals both with and without a wall are displayed in Figure 5b, using the combined dataset. It is evident from the graph that the InceptionV3 algorithm surpasses the others on the combined dataset. Specifically, when utilizing the InceptionV3 algorithm, a classification accuracy of 89.0% is achieved without a wall, whereas the same algorithm yields an 80.0% classification accuracy with the presence of a wall. The fusion of different deep learning models was tested which is illustrated in Fig. 6 . The highest accuracy values of 91% without the wall were achieved with feature fusion at the fully connected layers of VGG16 and InceptionV3 deep learning models shown in Fig. 5c. Furthermore, the highest accuracy of 83.33% was

achieved through the walls with the feature fusion of VGG16 and InceptionV3 deep learning models shown in Fig. 5d.

VII. Conclusion

In this work, an RF sensing-based head movement recognition system is proposed using Wi-Fi and radar, and state-of-the-art deep and machine learning algorithms. All directions of head movements were covered, such as Head up, Head down, Head left 90, Head right 90, Head left 45, and Head right 45. Wi-Fi data was passed to the InceptionV3 model and radar data to VGG16 models and the features of the two models were fused for the highest performance results of 91.85% without the walls and 83.33% accuracy was achieved through the walls. The proposed work is promising for many critical applications, such as fatigue detection and drowsiness for automated pilot monitoring systems and assistive car driving and alert systems, including wheel chair control for paralysis patients. Furthermore, the proposed system preserves the privacy concerns of users, which may exist in vision based systems.

Acknowledgment

This study received partial financial support from the Engineering and Physical Sciences Research Council (EPSRC) through grants: EP/T021063/1 (Q.H., M.I.), EP/T021020/1 (M.I.), and SAPHIRE2022 (Grant No: 2814).

References

[1] B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE transactions on multimedia*, vol. 17, no. 7, pp. 1107–1119, 2015.

[2] C. Zhao, C. Zheng, M. Zhao, Y. Tu, and J. Liu, "Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1859–1865, 2011.

[3] G. Borghini, G. Vecchiato, J. Toppi, L. Astolfi, A. Maglione, R. Isabella, C. Caltagirone, W. Kong, D. Wei, Z. Zhou, *et al.*, "Assessment of mental fatigue during car driving by using high resolution eeg activity and neurophysiologic indices," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6442–6445, IEEE, 2012.

[4] M. H. Alkinani, W. Z. Khan, and Q. Arshad, "Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent advances, requirements and open challenges," *Ieee Access*, vol. 8, pp. 105008–105030, 2020.

[5] X. Wang and C. Xu, "Driver drowsiness detection based on non-intrusive metrics considering individual specifics," *Accident Analysis & Prevention*, vol. 95, pp. 350–357, 2016.

[6] G. S. Larue, A. Rakotonirainy, and A. N. Pettitt, "Predicting reduced driver alertness on monotonous highways," *IEEE Pervasive Computing*, vol. 14, no. 2, pp. 78–85, 2015.

[7] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3017–3032, 2015.

[8] R. Safety-Statistics, "UK reported road casualties great britain, provisional results: 2021, department for transport." https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-provisional-results-2021. Accessed: 07 December 2022.

[9] S. Ansari, F. Naghdy, H. Du, and Y. N. Pahnwar, "Driver mental fatigue detection based on head posture using new modified relu-bilstm deep neural network," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[10] P. Chakraborty, M. A. Yousuf, and S. Rahman, "Predicting level of visual focus of human's attention using machine learning approaches," in *Proceedings of international conference on trends in computational and cognitive engineering*, pp. 683–694, Springer, 2021.

[11] T. Kujani and V. D. Kumar, "Head movements for behavior recognition from real time video based on deep learning convnet transfer learning," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15, 2021.

[12] A. E. Cetin, Y. Ozturk, O. Hanosh, and R. Ansari, "Review of signal processing applications of pyroelectric infrared (pir) sensors with a focus on respiration rate and heart rate detection," *Digital Signal Processing*, vol. 119, p. 103247, 2021.

[13] S. M. Preum, S. Munir, M. Ma, M. S. Yasar, D. J. Stone, R. Williams, H. Alemzadeh, and J. A. Stankovic, "A review of cognitive assistants for healthcare: Trends, prospects, and future directions," *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, pp. 1–37, 2021.

[14] H. Dsouza, J. Pastrana, J. Figueroa, I. Gonzalez-Afanador, B. M. Davila-Montero, and N. Sepúlveda, "Flexible, self-powered sensors for estimating human head kinematics relevant to concussions," *Scientific reports*, vol. 12, no. 1, pp. 1–8, 2022.

[15] S. Alghowinem, R. Goecke, M. Wagner, G. Parkerx, and M. Breakspear, "Head pose and movement analysis as an indicator of depression," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 283–288, IEEE, 2013.

[16] Y. Jiang, A. Sadeqi, E. L. Miller, and S. Sonkusale, "Head motion classification using thread-based sensor and machine learning algorithm," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[17] A. Al-Rahayfeh and M. Faezipour, "Eye tracking and head movement detection: A state-of-art survey," *IEEE journal of translational engineering in health and medicine*, vol. 1, pp. 2100212–2100212, 2013.

[18] K. S. Huang and M. M. Trivedi, "Robust real-time detection, tracking, and pose estimation of faces in video streams," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, pp. 965–968, IEEE, 2004.

[19] T. Horprasert, Y. Yacoob, and L. S. Davis, "An anthropometric shape model for estimating head orientation," in *3rd International Workshop on Visual Form, Capri, Italy*, 1997.

[20] E. N. A. Neto, R. M. Barreto, R. M. Duarte, J. P. Magalhaes, C. A. Bastos, T. I. Ren, and G. D. Cavalcanti, "Real-time head pose estimation for mobile devices," in *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 467–474, Springer, 2012.

[21] H. Hameed, M. Usman, A. Tahir, A. Hussain, H. Abbas, T. J. Cui, M. A. Imran, and Q. H. Abbasi, "Pushing the limits of remote rf sensing by reading lips under the face mask," *Nature Communications*, vol. 13, no. 1, pp. 1–9, 2022.

[22] H. Hameed, M. Usman, A. Tahir, K. Ahmad, A. Hussain, M. A. Imran, and Q. H. Abbasi, "Recognizing british sign language using deep learning: A contactless and privacy-preserving approach," *IEEE Transactions on Computational Social Systems*, 2022.

[23] S. A. Shah and F. Fioranelli, "Rf sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 26–44, 2019.

[24] S. Sun, A. P. Petropulu, and H. V. Poor, "Mimo radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98–117, 2020.

[25] C. Cheng, F. Ling, S. Guo, G. Cui, Q. Jian, C. Jia, and Q. Ran, "A real-time human activity recognition method for through-the-wall radar," in *2020 IEEE Radar Conference (RadarConf20)*, pp. 1–5, 2020.

[26] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "Tw-see: Human activity recognition through the wall with commodity wi-fi devices," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 306–319, 2019.