

Article

Enhancing CSI-Based Human Activity Recognition by Edge Detection Techniques

Hossein Shahverdi ¹, Mohammad Nabati ¹, Parisa Fard Moshiri ¹, Reza Asvadi ¹  and Seyed Ali Ghorashi ^{2,*} 

¹ Cognitive Telecommunication Research Group, Department of Electrical Engineering, Shahid Beheshti University G. C., Tehran 19839 69411, Iran; h.shahverdi@mail.sbu.ac.ir (H.S.); mo.nabati@mail.sbu.ac.ir (M.N.); p.fardmoshiri@mail.sbu.ac.ir (P.F.M.); r_asvadi@sbu.ac.ir (R.A.)

² Department of Computer Science & Digital Technologies, School of Architecture, Computing and Engineering, University of East London, London E16 2RD, UK

* Correspondence: s.a.ghorashi@uel.ac.uk

Abstract: Human Activity Recognition (HAR) has been a popular area of research in the Internet of Things (IoT) and Human–Computer Interaction (HCI) over the past decade. The objective of this field is to detect human activities through numeric or visual representations, and its applications include smart homes and buildings, action prediction, crowd counting, patient rehabilitation, and elderly monitoring. Traditionally, HAR has been performed through vision-based, sensor-based, or radar-based approaches. However, vision-based and sensor-based methods can be intrusive and raise privacy concerns, while radar-based methods require special hardware, making them more expensive. WiFi-based HAR is a cost-effective alternative, where WiFi access points serve as transmitters and users' smartphones serve as receivers. The HAR in this method is mainly performed using two wireless-channel metrics: Received Signal Strength Indicator (RSSI) and Channel State Information (CSI). CSI provides more stable and comprehensive information about the channel compared to RSSI. In this research, we used a convolutional neural network (CNN) as a classifier and applied edge-detection techniques as a preprocessing phase to improve the quality of activity detection. We used CSI data converted into RGB images and tested our methodology on three available CSI datasets. The results showed that the proposed method achieved better accuracy and faster training times than the simple RGB-represented data. In order to justify the effectiveness of our approach, we repeated the experiment by applying raw CSI data to long short-term memory (LSTM) and Bidirectional LSTM classifiers.

Keywords: human activity recognition; Internet of Things; deep learning; channel state information; convolutional neural networks



Citation: Shahverdi, H.; Nabati, M.; Fard Moshiri, P.; Asvadi, R.; Ghorashi, S.A. Enhancing CSI-Based Human Activity Recognition by Edge Detection Techniques. *Information* **2023**, *14*, 404. <https://doi.org/10.3390/info14070404>

Academic Editors: Vasco N. G. J. Soares, João M. L. P. Caldeira, Bruno Bogaz Zarpelão and Jaime Galán-Jiménez

Received: 13 June 2023

Revised: 5 July 2023

Accepted: 7 July 2023

Published: 14 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the past two decades, the Internet of Things (IoT) has emerged [1]. It refers to a group of computing devices and objects operating interrelatedly in a network to share and transfer data and information in a real-time, efficient, and fast manner without human intervention [2]. It covers intelligent homes, cities and networks, automation, AI, cybersecurity, telehealth, connected cars, hotel industries, and remote control [3]. One of the hot topics and emerging areas of research in the field of smart buildings and health monitoring which has been gaining considerable attention on both the academic and industrial sides, is Human Activity Recognition (HAR) [4]. HAR seeks to determine what specific daily activity is performed by users understanding the different responses that devices give each other because of the action [4,5].

Three different types of data can be collected for HAR, namely: vision-based, sensor-based, and radar-based [6] (As shown in Figure 1). Vision-based HAR focuses on visualizing data (images or videos) such as color, depth, and the skeleton 3D camera. In this method, there must be a direct sight to the users, and the presence of any hindrances and obstacles

would diminish the system's functionality and accuracy. In addition, vision-based HAR approaches are highly dependent on weather and lighting conditions, and due to continued monitoring, they can violate users' privacy [7]. The sensor-based approach is carried out either through different kinds of sensors assembled in a gadget, such as accelerometer, gyroscope, gravity, and orientation in smartphones, or wearable sensors, such as coats, shoes, smart watches/glasses, and gloves. Wearable sensors have achieved remarkable and acceptable results in HAR; however, they run into problems such as inconvenient use that may make them inefficient in some scenarios. Unlike the previously mentioned methods, radar-based HAR can be applied in scenarios where a direct line of sight is impossible between users and devices. Moreover, this approach is independent of environmental conditions such as light and weather [4]. However, equipping this approach is costly [8].

WiFi-based HAR, as a subgroup of radar-based approach, has gained considerable attention because of its advantages such as being less expensive, ubiquitously available, easy to deploy, power-efficient, and independent from light/weather characteristics [9]. In WiFi-based HAR, the aim is to find and highlight each specific activity's effect on the propagated signals, and based on this variation, predict the users' activities. Generally, WiFi-based HAR is carried out through two metrics of WiFi channels: Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) [9]. RSSI contains information about how the power of propagated signal varies through its way to the receiver and is ideally used in localization tasks.

The major concern with RSSI is that it is unstable, so it cannot capture dynamic changes during an activity performance [9,10]. In addition, the distance between the transmitter and receiver affects the accuracy of this method. This is while CSI is more stable and contains more information about wireless channels since it is measured at each orthogonal frequency-division multiplexing from each packet. In addition, it records both the amplitude and phase quantities; therefore, it can be implied how an activity would affect the propagated signal more accurately. Thus, it is preferred for WiFi-based HAR [11].

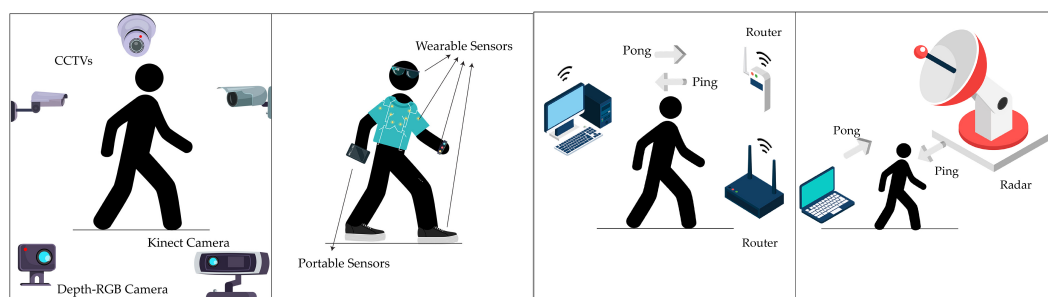


Figure 1. Different possible sensory data collections taken in HAR.

The problem with converting CSI data into an image is that the generated images in activities, such as walking and running, falling and lying down, have the same visual representations, which may make the classification difficult, in addition to applying different preprocessing techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to eliminate undesired values of noise. Based on our preliminary analysis and observations, we hypothesize that by applying edge detection techniques to the CSI images in WiFi-based HAR and utilizing the Canny, Prewitt, Sobel, and Laplacian of Gaussian (LoG) filters, we can improve the accuracy and overcome the challenge of visual representations of similar activities, leading to more effective activity recognition compared to existing methods. We applied edge detection techniques and used mentioned well-known filters on the converted RGB-based images. After applying edge detector filters, we repeated the classification process by the same CNN layers. We witnessed an increase in accuracy and elimination of overfitting in all the studied datasets. We analyzed three publicly available WiFi datasets gathered by [12–14], and based on their characteristics, we applied edge detection techniques to the CSI images

to extract more features out of them. In order to show the effectiveness of our proposed method, we compared the results of our proposed method with two networks: LSTM and Bidirectional-LSTM Deep Learning (DL) model proposed in [12,13].

The rest of this paper is organized as follows: a brief review of the previous literature with different modalities and What is Channel State Information (CSI) is presented in Section 2. In Section 3, we present previously studied papers in the field of WiFi-based HAR with each drawback alongside the different approach we have taken in this study. The overview of the analyzed datasets and the proposed method is described in Section 4. The proposed activity recognition algorithm, experimental evaluation, and results are presented in Section 5. Finally, conclusion and future works are described in Section 6.

2. Background

Human Activity Recognition (HAR) is a research field that classifies human actions using sensor data. It is used in computer vision and Machine Learning (ML), aiming to detect and recognize human activities automatically. HAR has potential in various sectors, including healthcare, sports, and security, but requires ethical and transparent execution to meet privacy concern. Sensor modalities used in HAR include:

- RGB, Skeleton, and Depth data: RGB data provides visual information about humans, skeleton data focuses on the spatial relationships between body parts, and depth data provides information about the 3D structure of a scene [15].
- Audio Data: Captures human activities based on audio patterns, useful when visual information is limited [16].
- Wearable Sensors: Capture acceleration data, measuring changes in movement and velocity, useful for detecting activities involving body motion [17].
- Radar Data: Represents reflected signals, used to detect human presence and movement [18].
- WiFi Signal Data: Utilizes wireless signals to recognize activities based on changes caused by human motion. The choice of sensor modality depends on the application's specific requirements, such as environmental conditions and available hardware resources [12].

WiFi signals, which are commonly used for wireless communication, can also be utilized to capture human activities within their range [12]. Channel State Information (CSI) refers to the information carried by WiFi signals about the wireless channel between the transmitter (WiFi access point) and the receiver (WiFi device) [12,14]. It includes data about the amplitude, phase, and frequency response of the signals. When a human engages in activity, their movements introduce changes in the wireless channel characteristics. These changes affect the properties of the WiFi signals, which can be captured and analyzed using CSI. By examining the variations in the amplitude and phase of the WiFi signals, we can infer specific characteristics of human activities. To obtain CSI, specialized WiFi devices with multiple antennas are used. These devices have the capability to measure the variations in the wireless channel caused by human movements. The measurements are typically collected at a high temporal resolution, capturing the rapid changes in the wireless signals [12,13]. Suppose we have a MIMO communication system has several subcarriers in a single connection between a transmitter and receiver. Each subcarrier has unique CSI values. If we assume there are t transmitters, and r receivers operating in a MIMO system, CSI for the m^{th} packet can be represented as a matrix of

$$CSI_m = \begin{pmatrix} H_{1,1} & \cdots & H_{1,r} \\ \vdots & \ddots & \vdots \\ H_{t,1} & \cdots & H_{t,r} \end{pmatrix} \quad (1)$$

where $H_{t,r}$ is a complex-valued vector for each subcarrier. Each element of this matrix contains information about the amplitude and phase of the propagated signal from the

t^{th} transmitter and r^{th} receiver. A single element of CSI at central frequency of f_k can be described as:

$$H_{t,r}(f_k) = ||H_{t,r}(f_k)|| \times e^{i\angle H_{t,r}(f_k)} \quad (2)$$

where $||H_{t,r}(f_k)||$ and $\angle H_{t,r}(f_k)$ represents the amplitude and the phase, respectively. In addition, the n^{th} received sequence at time t in the subcarrier with the central frequency of f_k can be shown as:

$$R(f_k, t) = T(f_k, t) \times CSI + W \quad (3)$$

where $T(f_k, t)$ and $R(f_k, t)$ are the transmitted and received sequence, respectively, whereas W stands for the environmental noise in the experiment [13].

3. Related Works

In this section, we will focus on previous studies solely in the era of WiFi-based HAR and finally, we will explain how our approach is different compared to former ones.

Wang et al. [19] proposed CARM (Channel State Information-based Human Activity Recognition and Monitoring), a system that utilized WiFi signals to accurately recognize and monitor human activities. The system addressed the limitations of existing WiFi-based recognition systems by introducing the CSI-speed model and the CSI-activity model. The CSI-speed model established the relationship between CSI power variations and movement speeds, while the CSI-activity model represented the movement speeds of body parts to specific activities. CARM offered several advantages. It enabled precise movement feature extraction using commercial WiFi devices, allowing for quantitative assessment of activities. The model-based approach guided system design and noise removal techniques, leveraging insights such as the frequency range of CSI variations caused by human activities. The system handled challenges such as noisy CSI values by employing PCA-based denoising and addressed variations in activity performance through the use of Hidden Markov Models. Robustness in different environments is achieved by performing data fusion from multiple WiFi links. Experimental results demonstrate high recognition accuracy, averaging 96% on a comprehensive activity database. CARM's implementation on commercial WiFi devices, such as routers and laptops, showcased its practical applicability. Although this study could achieve an acceptable performance it had some drawbacks. For example, the reliance on specific commercial WiFi devices, such as Intel 5300 WiFi cards, may limit the system's compatibility with other devices. This hardware dependency could hinder its adoption in different settings or with future advancements in WiFi technology.

Ding et al. [20] presented a WiFi Channel State Information (CSI)-based human activity recognition approach called HARNN (Human Activity Recognition using Deep Recurrent Neural Network). The system addressed the need for accurate activity recognition using commercial WiFi devices in applications such as smart homes and interactive games. HARNN incorporates four key techniques to recognize different human activities. Firstly, it utilized a novel two-level decision tree that efficiently uses variance and correlation coefficients of raw WiFi CSI data. A linear regression method is employed to determine the optimal parameters for the decision tree, reducing false activity detection caused by noise in indoor environments. To overcome the challenge of random noise interference, a noise removal mechanism based on discrete wavelet transform (DWT) is introduced. This strategy effectively filters out noise while preserving WiFi signal details, enhancing the accuracy of activity recognition. Two representative features, channel power variation (CPV) in the time domain and time-frequency analysis (TFA) in the frequency domain, were extracted from denoised WiFi CSI data. These features provided profound characterization of various human activities. The proposed HARNN employed a recurrent neural network (RNN) model, specifically with a long short-term memory (LSTM) block. Experiments conducted on commercial WiFi devices validate the high performance of HARNN in human activity recognition using WiFi CSI. The system outperforms benchmark approaches in terms of recognition accuracy. The experiments also demonstrate the robustness of HARNN in different indoor environments.

Yuan et al. [21] proposed a CSI-based device-free HAR (CDHAR) system that integrates WiFi-sensing radar on UAVs for human activity recognition. The system addressed two disadvantages of existing CSI-based HAR systems: manually setting detection thresholds and the use of a sole classifier for recognition. To overcome these challenges, CDHAR employs machine learning and kernel density estimation (KDE) to obtain adaptive detection thresholds, improving adaptability and instantaneity in different wireless environments. Additionally, a random subspace classifier ensemble method is introduced, using frequency domain features instead of time domain features, and ensuring higher recognition accuracy compared to existing systems. The advantages of CDHAR include the adaptive detection threshold algorithm, which accurately extracts activity durations even in varying wireless environments. The use of frequency domain features and the random subspace classifier ensemble method further enhances recognition accuracy. However, there are some drawbacks to consider. Firstly, the reliance on machine learning and KDE for adaptive detection thresholds introduces additional complexity and computational overhead. Secondly, focusing primarily on frequency domain features may neglect important temporal patterns and nuances present in the time domain.

Arshad et al. [22] present Wi-Chase, a sensorless system for human activity detection using Channel State Information (CSI) from WiFi packets. The system aims to overcome limitations of existing approaches by fully utilizing all available subcarriers of the WiFi signal and incorporating variations in both phases and magnitudes. Wi-Chase introduces an adaptive Activity Detection Algorithm (ADA) that evaluates the variations in all subcarriers to improve recognition accuracy by leveraging detailed correlated information content. The system employs subcarrier level majority voting and utilizes both amplitude and phase features, resulting in higher classification accuracy compared to previous works. The authors construct a diverse dataset of activities from different users and analyze the system's performance with varying numbers of subcarriers and communication links. Experimental results demonstrate that Wi-Chase achieves an average accuracy greater than 97% for multiple communication links. Overall, Wi-Chase represents a novel sensorless system that effectively detects and classifies human activities using WiFi CSI, providing a promising approach for activity recognition in real-world scenarios.

In our research, we built upon the WiFi-based HAR concept and employed a Convolutional Neural Network (CNN) as a classifier. As a preprocessing step, we applied edge-detection techniques to improve the quality of activity detection. Instead of directly using the CSI data, we converted it into RGB images and used them as inputs to the CNN. This conversion allowed us to leverage the powerful image-processing capabilities of CNNs. We conducted experiments using three available CSI datasets which were collected via different devices and modalities and compared the performance of our method with a simple RGB-represented data approach. The results of our experiments demonstrated that our proposed method achieved better accuracy and faster training times compared to the simple RGB-represented data approach. To further justify the effectiveness of our approach, we also repeated the experiments by applying raw CSI data to Long Short-Term Memory (LSTM) and Bidirectional LSTM classifiers. These additional experiments provided further evidence of the superiority of our method in capturing and utilizing the channel information for accurate activity recognition. By focusing on WiFi-based HAR and leveraging the rich information provided by CSI, our approach offers distinct advantages over traditional vision-based, sensor-based, and radar-based methods. It provides a cost-effective, non-intrusive, and privacy-friendly solution for human activity recognition. Our research contributes to the advancement of WiFi-based HAR techniques and demonstrates the potential for improved accuracy and efficiency by incorporating advanced machine learning algorithms and preprocessing techniques.

4. System Model

4.1. CSI Dataset and the Collection Methodology

In this section, we will discuss the specialized hardware and software used for CSI data collection and explore the datasets we used for this study. According to the previous studies, the most prominent hardware to extract CSI data is Intel 5300 WiFi Network Interface Card (NIC) or Linux 802.11n, Atheros CSI tools such as AR9580, AR9590, AR9344, and QCA9558, and Raspberry Pi or Nexmon CSI Tool [23]. Since Nexmon CSI Tool offers better and more promising features such as enabling access and monitoring MIMO antennas up to 44 numbers and also collecting the information on a channel between a specific transmitter and receiver, and additionally, due to its easy deployment on Raspberry Pi 3B+, Pi 4B, Google Nexus 5, and Routers, and more importantly its cost-friendly feature, recent studies have been focusing on collecting data using this approach [12]. In this study, we analyzed three available CSI-based datasets gathered by different researchers. The first dataset, collected by Moshiri et al. [12], involved the use of Raspberry Pi 4 and Tp-Link Archer C20 as an access point. The experiments were conducted in a bedroom environment, where three subjects in the age range of 25 to 70 performed seven different activities, including bend, fall, stand up, sit down, lie down, run, and walk. Each activity was performed for a duration of 10 to 20 s. The data collection was done in a 20 MHz bandwidth on channel 36 in the IEEE 802.11ac standard with 52 subcarriers. The second dataset, collected by Schäfer et al. [13], utilized a dual-band router Fritzbox operating at the 5 GHz frequency band. Raspberry Pi 4B was used as the data collection device. One CSI Tool was employed to extract CSI data at an 80 MHz bandwidth with 128 subcarriers. This dataset consisted of four activities: empty, stand up, sit down, lie, and walk. The third dataset, collected by Ashleibta et al. [14], involved the use of two Universal Software Radio Peripheral (USRP) devices. One device served as the transmitter, and the other as the receiver, operating at 5GHz with 52 subcarriers. Ubuntu virtual machine and Gnu radio were used to generate data traffic. The dataset contained categories such as empty, sitting, standing, and walking. Table 1 in our paper summarizes the key features of these datasets, including the hardware and software used, frequency bands, bandwidths, subcarriers, and the activities performed during data collection. By utilizing these publicly-available datasets, we ensured the diversity and generalizability of our proposed method across different scenarios and activities.

Table 1. Summary of Studied Datasets.

Dataset	Tool Used to Collect	Bandwidth & Number of Subcarriers (Including Zero & Pilot)	Number of Activities
Schäfer et al. [13]	Raspberry Pi 4B + Nexmon CSI Tool	80 MHz and 256 subcarriers 802.11ac Standard	1 + 4: Empty, Standup, Sitdown, Walk, Lie down (in total 1800 number)
Ashleibta et al. [14]	Universal Software Radio Peripheral devices	3.75 GHz and 52 Subcarriers	1 + 3: Empty, Sitting, Standing, Walking (in total 540 number)
Moshiri et al. [12]	Raspberry Pi 4B Nexmon CSI Tool	40 MHz and 52 Subcarriers 802.11ac Standard	7: Bend, Walking, Running, Standing up, Sitting down, Falling, Lying down (in total 420 number)

4.2. Neural Networks

Every specific action has a different effect on CSI, enabling the system to detect, recognize, or even predict the action [12]. Many previous studies took the advantages of ML and DL algorithms such as SVM, NB, Hidden Markov Model (HMM), RF, LSTM and BiLSTM [13], 1D and 2D CNN [12]. Although these algorithms have scored good accuracies in the classification task, they face problems such as overfitting, small batch sizes, and weight decay due to the small size of datasets. In this research, we took advantage of the DL model due to its superiority over ML algorithms, such as the capability of feature

engineering on its own. We presented two custom DL models based on 2D-CNN with fewer computational complexity and training time for classification tasks.

CNN

CNN is a feed-forward DL algorithm that can take images as input and assign weights and biases to the various objects in images to make them distinguishable from each other [12]. Compared to the other classification algorithms, CNN requires less preprocessing, making it more efficient for a situation where a decision must be made fast. There are various components called layers present in a simple CNN architecture. CNN uses filters (or kernel) to extract an image's spatial and temporal feature dependencies. Some of the most important layers are Convolution, Padding, Pooling, Dropout, Batch normalization, and fully connected. The objective of initial Convolution layers is to extract low-level features such as color and gradient variation by a convolution operation. They can understand high-level features such as edges, lines, and curves by increasing the number. In order to save the information on the corner of images, this layer is used to preserve that information by adding extra columns and rows on the outer dimensions of images. Like the convolution layer, the pooling layer is responsible for reducing the spatial size of the convolved features, which is called dimensionality reduction [12]. This procedure is maintained to decrease the computational efforts required to process the data by highlighting dominant features and taking out trifling ones. As mentioned above, training these data may result in a phenomenon called "Overfitting" when we do not have a plethora of data. A remedy to overcome this issue is applying a dropout layer after each convolution layer, which randomly sets some input weights to 0 with a frequency rate during the training process. The other frequently-used layer aims to decrease the training time by normalizing the layers' input, such as re-scaling and re-centering. Finally, the fully connected or Dense layer is used to learn non-linear high-level features as represented by the output of the convolution layer. Every neuron is connected to the previous and subsequent layers in this layer. Based on the complexity of the model, the number of layers varies, and specific layers are used.

5. Proposed Methodology and Experimental Evaluations

Image analysis is the technique to extract features automatically [24]. Some of the most common techniques are image segmentation, texture and motion analysis, and edge extraction and detection [24]. The edge detection process significantly reduces the amount of processed data by filtering unnecessary information. Generally, the edge detection methods are categorized into two sections: 1—Gradient and 2—Laplacian. The gradient method searches for minimum and maximum values in the first derivative of the image while Laplacian for the zero crossings in their second derivative. This paper applies edge extraction techniques to our generated CSI images and use four well-known filters, including Sobel, Canny, Prewitt, and Laplacian of Gaussian (LoG), in which an improvement was seen both in terms of accuracy and consumed time for each training and testing phase. The main goals to be achieved by using edge detection techniques are:

- Detect edges with least probability of error;
- Mitigate the amount of the noise presented in images in order to prevent false edges.

In the following subsection, we introduce some of the well-known filters briefly.

5.1. Sobel Filter

The Sobel filter is a subgroup of gradient operators used to clarify the local transformation, such as sharp edges. A Sobel edge detector comprises a pair convolution layer in which the second one is a transformed version of the former one. These two layers are as follows:

$$S_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} \times A \quad (4)$$

and

$$S_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \times A \quad (5)$$

where A is the source image. P_x responds to edges through x -axis while P_y to ones in the y -axis. In order to find the absolute magnitude and the direction of the gradient at each point, these two metrics can be combined as follows:

$$P = \sqrt{P_x^2 + P_y^2} \quad (6)$$

$$\Theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (7)$$

Figure 2 represents the original and filtered images using the Sobel filter:

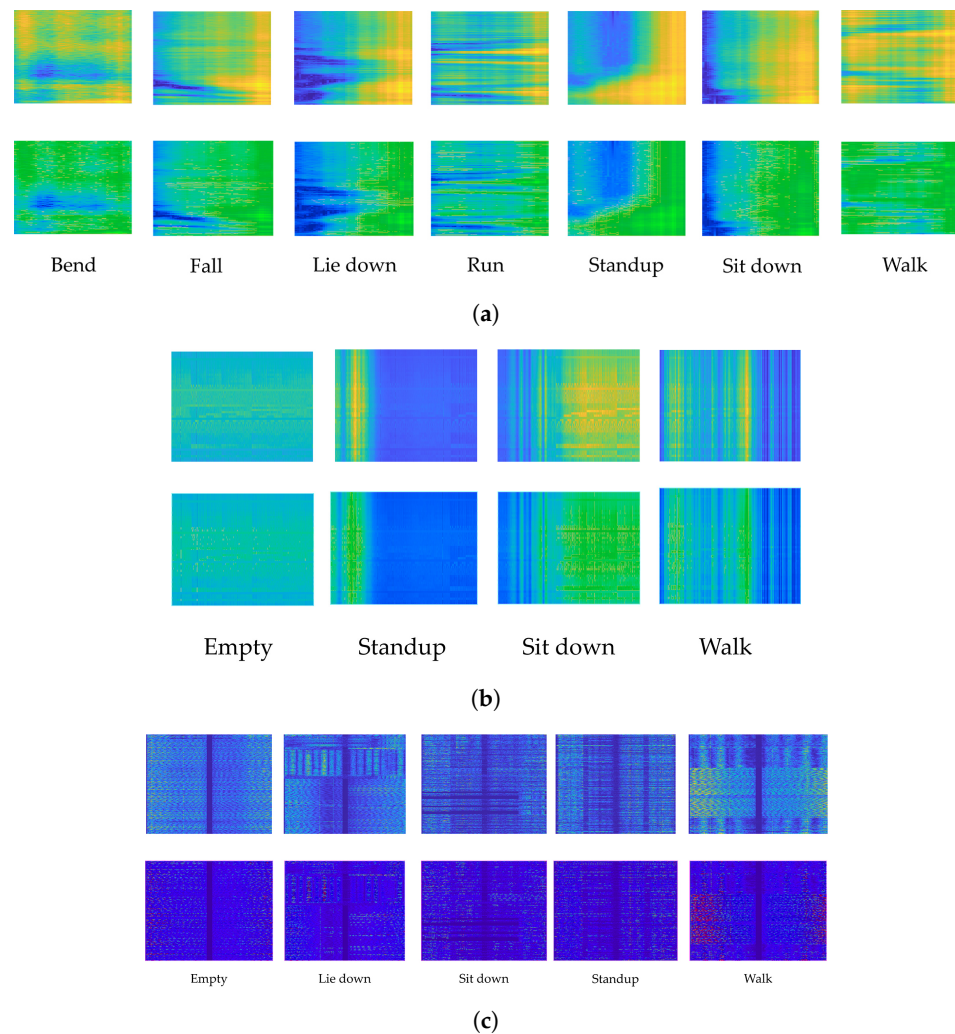


Figure 2. Original and filtered images using Sobel edge detector for: (a) Moshiri et al. [12] (b) Ashleibta et al. [14] and (c) Schäfer et al. [13].

5.2. Canny Filter

Canny is one of the well-studied and applied filters due to its features. The main contribution of the canny filter are as follow: (1) Apply a Gaussian filter to minimize the amount of presented noise. (2) Verify if the detected edges are either true with high

probability or misrepresented by applying a threshold. The most common 5×5 Gaussian filter, which is also used by default in our experiment, is:

$$C = \frac{1}{159} \times \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \times A \quad (8)$$

5.3. Prewitt Filter

The Prewitt filter is another filter that was used in this research. Like the Sobel filter, it finds edges by convolving two masks on the horizontal and vertical axes. The vertical and horizontal mask used in this filter is as follows:

$$P_x = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix} \times A \quad (9)$$

$$P_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix} \times A \quad (10)$$

P_x prominent vertical edges while P_y on the horizontal ones. These two matrices work as a first-order derivative and calculate the difference of pixel intensities in an edge region. As the center column is zero, it does not include the original values of an image. However, rather it calculates the difference between right and left pixel values around that edge, increases the edge intensity, and it becomes enhanced comparatively to the original image.

5.4. LoG Filter

Laplacian of Gaussian (LoG) filter is considered the second derivative operation and comprises two parts: Laplacian and Gaussian filter. The Laplacian of an image highlights regions of rapid intensity change and, therefore, is often used for edge detection. The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce its sensitivity to noise. Hence, the two variants will be described together here.

5.5. Proposed Method

This subsection will describe the steps taken in this paper to detect daily activities from CSI data from three available datasets. Based on the dataset's characteristics, preprocessing techniques such as Principal Component Analysis (PCA), Normalization, and Linear Discriminant Analysis (LDA) were used for dimensionality reduction and denoising purposes. Data from [13] consist of much noise. In order to decrease these unwanted values, we applied both PCA with three components and LDA. After applying preprocessed algorithms on the CSI data, data were converted into RGB-based images, followed by different edge detector filters to highlight more features from the generated images. For classification, we proposed a CNN-based neural network, shown in Figure 3. The model comprises batch normalization layer which was implemented after the dense layer to reduce the probability of overfitting. In order to prevent overfitting, dropout layers were implemented after each convolution and dense layer. First, original images are fed into CNN models, and filtered ones are set to be fed in the next step. Figure 4. describes the roadmap we took in this paper and Figure 5 shows the model architecture.

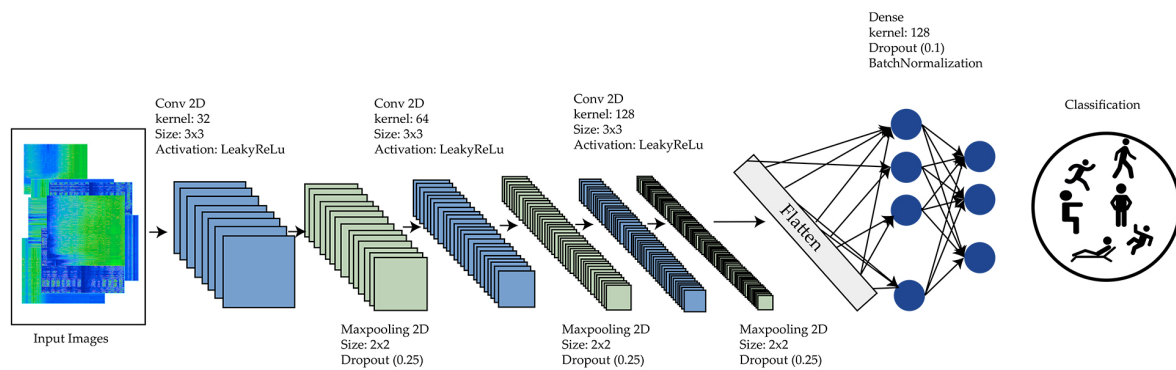


Figure 3. The proposed CNN model.

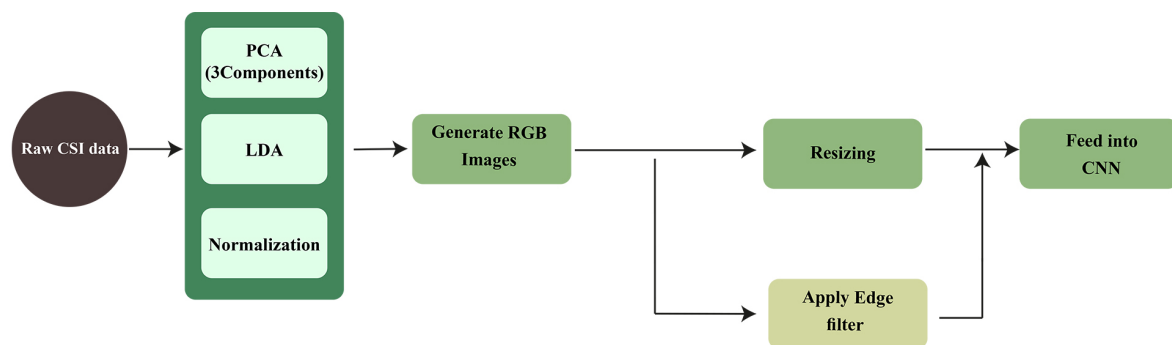


Figure 4. The roadmap taken in this paper.

5.6. Experimental Setups and Results

Tensorflow 2.8 with Python 3.8 was used for simulation purposes, and accelerated by GeForce RTX 3060. Since the data for [13] were noisy, we first applied PCA with five components and LDA to mitigate these unwanted factors. In order to make the numerical values of the scrutinized datasets to be in the range of RGB intensity pixels, all of them were normalized between 0 to 255, followed by converting them into an RGB-based representation. Since the number of data was not considerable, we adopted K-fold cross-validation with $k = 5$ and in order to prevent overfitting, batch normalization and dropout layers were added to the proposed CNN architectures. For the first part of the experiment, generated images are fed into two proposed CNNs. Despite adopting techniques to prevent overfitting, we faced this issue in all three dataset classifications. For the second scenario, an image preprocessing technique called edge detection was applied using four different filters on the generated images. Then, we repeated the classification using the same CNN architectures, in which not only we see an increase in the accuracy for all evaluated datasets, but also witnessed that the overfitting problem has been solved. Figure 6 demonstrates the achieved accuracies for all evaluated datasets using four edge detection filters. Further, in order to legitimize our proposed method, we compared our results with two neural networks called Long Short-Term Memory (LSTM) and Bidirectional Long-Short Term Memory (BiLSTM), proposed by [12,13], respectively, which the results are demonstrated in Figures 7–9 for [12–14], respectively. As we witnessed, our proposed method outperformed the other two methods, both in system performance and consumed time for training and test. Table 2, shows the consumed time for training and test phases.

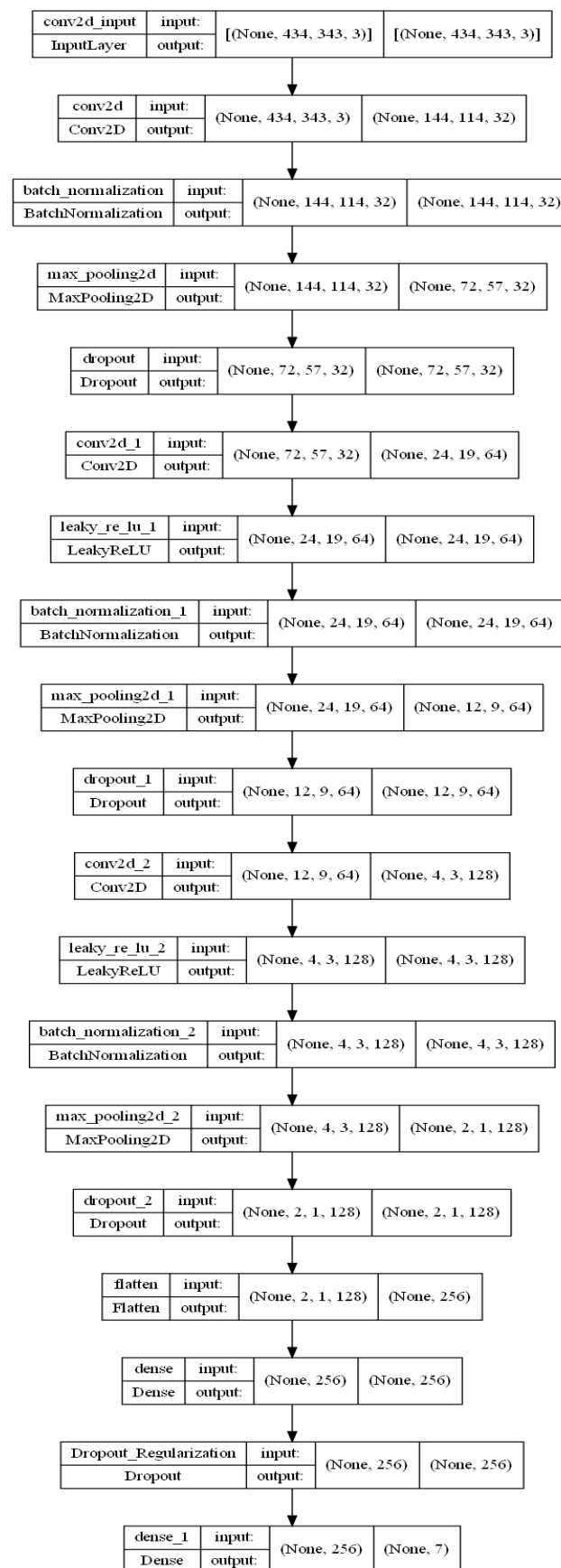


Figure 5. Proposed CNN architecture for classification.

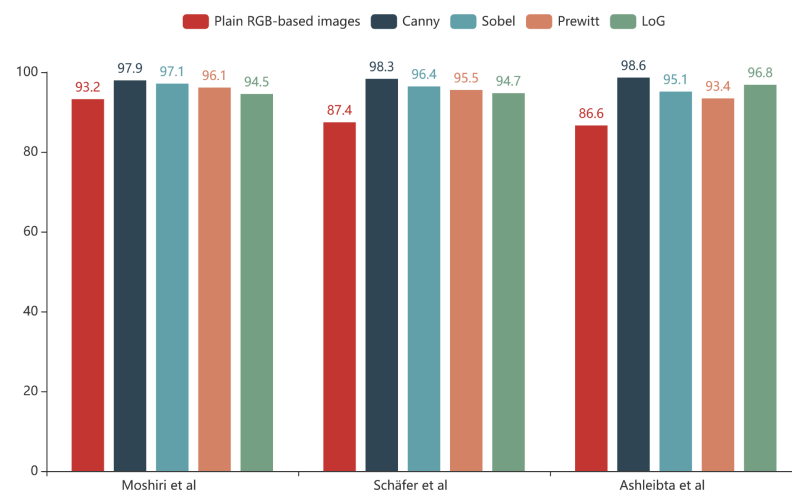


Figure 6. The accuracies, before and after applying different edge filters for two proposed CNN architectures [12–14].

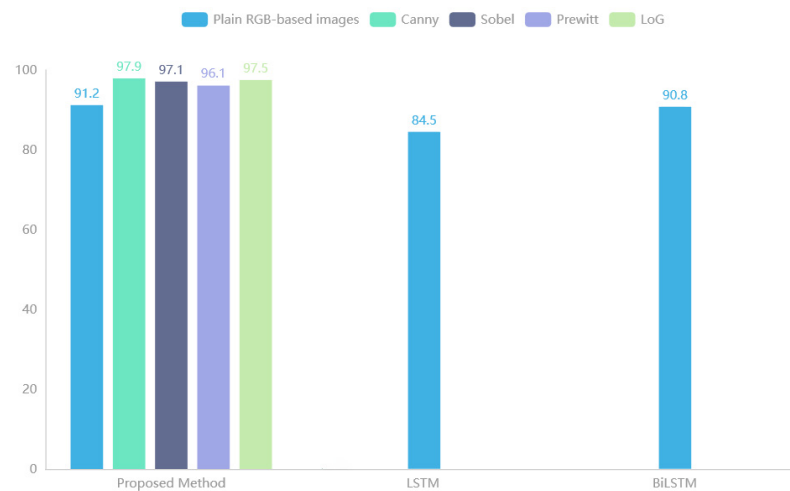


Figure 7. The accuracies acquired for Moshiri et al. [12] using three different modalities: (1)—Proposed Method. (2)—LSTM. (3)—BiLSTM.

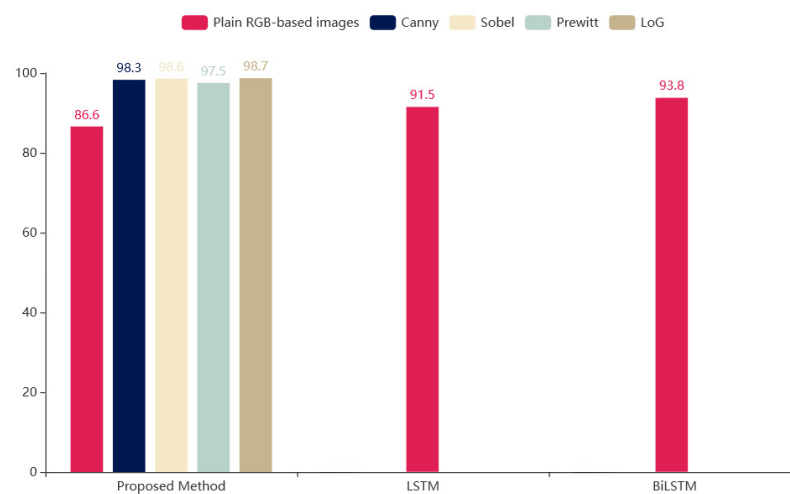


Figure 8. The accuracies acquired for Schäfer et al. [13] using three different modalities: (1)—Proposed Method. (2)—LSTM. (3)—BiLSTM.

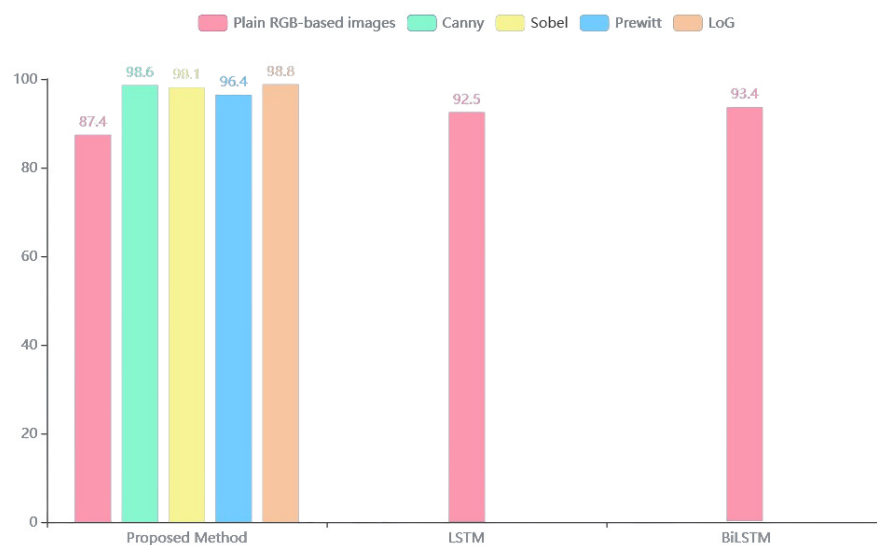


Figure 9. The accuracies acquired for Ashleibta et al. [14] using three different modalities: (1)—Proposed Method. (2)—LSTM. (3)—BiLSTM.

Table 2. Consumed time on training and testing phase for each epoch in milliseconds.

Time (In Milliseconds)	Proposed Method (In Average)	BiLSTM	LSTM
Training	15	17	25
Testing	5	8	14

6. Conclusions and Future Works

WiFi-based Human Activity Recognition (HAR), due to its specific characteristics such as ease of deployment, cost-effectiveness, ubiquitousness, and preserving the privacy of users has gained attention both on the academic and industrial applications. In this paper, we studied three publicly available CSI-based datasets. We converted these data into an RGB-image representation. In order to extract more features from generated images, we utilized an image preprocessing technique called edge detection and applied the most four well-known and used filters, namely, Canny, Sobel, Prewitt, and LoG to our images, following which we finally classified using a Neural Network method called 2D Convolutional Neural Networks. The results show improvement both in terms of consumed training time and system overall performance. We also duplicated the proposed methods described in [12,13], called Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM), in order to compare the performance of our proposed method and witnessed that our proposed method, meaning applying edge detection filters on generated CSI images, has better performance and consumes less time in training phase. While our research has yielded promising results, it is essential to acknowledge certain limitations. Although the implementation of edge detection filters enhanced the classification performance of our CNN model, the choice of specific filters (Canny, Sobel, Prewitt, and LoG) was based on their popularity and well-known properties. Further exploration of alternative or advanced edge detection techniques could potentially enhance the accuracy and robustness of activity detection. In addition, based upon our perspective, there are some branches in the field of WiFi-based HAR which could be potential directions for future work. For example, undertaking field experiments in real-world settings, such as smart homes or healthcare facilities, would provide insights into the practical challenges and performance of WiFi-based HAR systems. Addressing issues related to environmental variations, interference, and user privacy concerns in real-world deployment scenarios would be crucial for advancing the adoption of these systems. Another potential case study would be related to feature engineering which means investigating additional feature extraction techniques beyond edge detection could further enhance the discriminative power of CSI images for

activity recognition. Techniques such as deep feature learning or transfer learning from large-scale datasets could be investigated to enhance the representation and interpretability of the extracted features.

Author Contributions: Conceptualization, M.N. and S.A.G.; Methodology, M.N. and S.A.G.; Software, H.S.; Resources, P.F.M. and R.A.; Writing—original draft, H.S.; Writing—review & editing, P.F.M., R.A. and S.A.G.; Supervision, S.A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in GitHub: <https://github.com/parisafm/CSI-HAR-Dataset> (accessed on 27 October 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hassan, Q.F. *Internet of Things A to Z: Technologies and Applications*, 1st ed.; Wiley-IEEE Press: Hoboken, NJ, USA, 2018; pp. 5–45, ISBN 978-1-119-45674-2
- Dey, N.; Hassanien, A.E.; Bhatt, C.; Ashour, A.S.; Satapathy, S.C. *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 30, pp. 199–243, ISBN 978-3-319-86864-6.
- Perera, C.; Liu, C. H.; Jayawardena, S. The Emerging Internet of Things Marketplace from an Industrial Perspective: A Survey. *IEEE Trans. Emerg. Top. Comput.* **2015**, *3*, 585–598. [\[CrossRef\]](#)
- Wang, F.; Feng, J.; Zhao, Y.; Zhang, X.; Zhang, S.; Han, J. Joint Activity Recognition and Indoor Localization with WiFi Fingerprints. *IEEE Access* **2019**, *7*, 80058–80068. [\[CrossRef\]](#)
- Vlachostergiou, A.; Stratogiannis, G.; Caridakis, G.; Siolas, G.; Mylonas, P. *Smart Home Context Awareness Based on Smart and Innovative Cities*; Association for Computing Machinery: New York, NY, USA, 2015; ISBN 9781450335805.
- Palipana, S.; Rojas, D.; Agrawal, P.; Pesch, D. FallDeFi: Ubiquitous Fall Detection using Commodity WiFi Devices. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **2018**, *1*, 155. [\[CrossRef\]](#)
- Moshiri, P.F.; Navidan, H.; Shahbazian, R.; Ghorashi, S.A.; Windridge, D. Using GAN to Enhance the Accuracy of Indoor Human Activity Recognition. In Proceedings of the 10th Conference on Information and Knowledge Technology, Tehran, Iran, 31 December 2019.
- Ahad, M.A.R.; Ngo, T.T.; Antar, A.D.; Ahmed, M.; Hossain, T.; Muramatsu, D.; Makihara, Y.; Inoue, S.; Yagi, Y. Wearable Sensor-Based Gait Analysis for Age and Gender Estimation. *Sensors* **2020**, *20*, 2424. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nabati, M.; Ghorashi, S.A.; Shahbazian, R. Joint Coordinate Optimization in Fingerprint-Based Indoor Positioning. *IEEE Commun. Lett.* **2021**, *25*, 1192–1195. [\[CrossRef\]](#)
- Zhang, W.; Zhou, S.; Yang, L.; Ou, L.; Xiao, Z. WiFiMap+: High-Level Indoor Semantic Inference with WiFi Human Activity and Environment. *IEEE Trans. Veh. Technol.* **2019**, *68*, 7890–7903. [\[CrossRef\]](#)
- Chen, Z.; Zhang, L.; Jiang, C.; Cao, Z.; Cui, W. WiFi CSI based passive Human Activity Recognition Using Attention Based BLSTM. *IEEE Trans. Mob. Comput.* **2019**, *18*, 2714–2724. [\[CrossRef\]](#)
- Fard Moshiri, P.; Shahbazian, R.; Nabati, M.; Ghorashi, A. A CSI-based human activity recognition using Deep Learning. *Sensors* **2021**, *21*, 7225. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schäfer, J.; Barrsiwal, B.; Kokhkhharova, M.; Adil, H.; Liebehenschel, J. Human Activity Recognition Using CSI Information with Nexmon. *Sensors* **2021**, *11*, 8860. [\[CrossRef\]](#)
- Ashleibta, A.M.; Taha, A.; Khan, M.A.; Taylor, W.; Ahsen, T.; Zoha, A.; Abbasi, Q.; Imran, M.A. 5G-enabled contactless multi-user presence and activity detection for independent assisted living. *Sci. Rep.* **2021**, *11*, 17590. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bagate, A.; Shah, M.A. Human Activity Recognition using RGB-D Sensors. In Proceedings of the International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 15–17 May 2019. [\[CrossRef\]](#)
- Reynolds, F.; Neto, C.; Machado, J. Deep Learning for Activity Recognition Using Audio and Video. *Electronics* **2022**, *11*, 782. [\[CrossRef\]](#)
- Uddin, M.Z.; Soylu, A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. *Sci. Rep.* **2021**, *11*, 16455. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, Z.; Kernec, J.L.; Abbasi, Q.; Fioranelli, F.; Yang, S.; Romain, O. Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms. *Sci. Rep.* **2023**, *13*, 3473. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, W.; Liu, A.X.; Shahzad, M.; Ling, K.; Lu, S. Device-Free Human Activity Recognition Using Commercial WiFi Devices. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1118–1131. [\[CrossRef\]](#)
- Ding, J.; Wang, Y. WiFi CSI-Based Human Activity Recognition Using Deep Recurrent Neural Network. *IEEE J. Mag.* **2019**, *7*, 174257–174269. [\[CrossRef\]](#)

21. Yuan, H.; Yang, X.; He, A.; Li, Z.; Zhang, Z.; Tian, Z. Features extraction and analysis for device-free human activity recognition based on channel state information in 5G wireless communications. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 36. [CrossRef]
22. Arshad, S.; Feng, C.; Liu, Y.; Hu, Y.; Yu, R.; Zhou, S.; Li, H. Wi-chase: A WiFi based human activity recognition system for sensorless environments. In Proceedings of the IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), Macau, China, 12–15 June 2017.
23. Raspberry Pi Hardware Reference. 2014. Available online: <https://www.raspberrypi.com/> (accessed on 30 October 2022).
24. Reeves, S.J. Image restoration: Fundamentals of image restoration. *Acad. Press Libr. Signal Process.* **2014**, *4*, 165–192. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.