Review

Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances

Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng and Nabil Alshurafa

Special Issue
Deep Learning Methods for Human Activity Recognition and Emotion Detection

Edited by
Prof. Dr. Mario Munoz-Organero

Review

Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances

Shibo Zhang [1,2,*,†], Yaxuan Li [3,†], Shen Zhang [4], Farzad Shahabi [1,2], Stephen Xia [5], Yu Deng [6] and Nabil Alshurafa [1,2]

1 Department of Computer Science, McCormick School of Engineering, Northwestern University, Mudd Hall, 2233 Tech Drive, Evanston, IL 60208, USA; farzadshahabi2024@u.northwestern.edu (F.S.); nabil@northwestern.edu (N.A.)
2 Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, 680 N. Lakeshore Dr., Suite 1400, Chicago, IL 60611, USA
3 Electrical and Computer Engineering Department, McGill University, McConnell Engineering Building, 3480 Rue University, Montréal, QC H3A 0E9, Canada; yaxuanli123@gmail.com
4 School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive, Atlanta, GA 30332, USA; shenzhang@gatech.edu
5 Department of Electrical Engineering, Columbia University, Mudd 1310, 500 W. 120th Street, New York, NY 10027, USA; sx2194@columbia.edu
6 Center for Health Information Partnerships, Feinberg School of Medicine, Northwestern University, 625 N Michigan Ave, Chicago, IL 60611, USA; yudeng2015@u.northwestern.edu
* Correspondence: shibozhang2015@u.northwestern.edu
† These authors contributed equally to this work.

Abstract: Mobile and wearable devices have enabled numerous applications, including activity tracking, wellness monitoring, and human–computer interaction, that measure and improve our daily lives. Many of these applications are made possible by leveraging the rich collection of low-power sensors found in many mobile and wearable devices to perform human activity recognition (HAR). Recently, deep learning has greatly pushed the boundaries of HAR on mobile and wearable devices. This paper systematically categorizes and summarizes existing work that introduces deep learning methods for wearables-based HAR and provides a comprehensive analysis of the current advancements, developing trends, and major challenges. We also present cutting-edge frontiers and future directions for deep learning-based HAR.

Keywords: review; human activity recognition; deep learning; wearable sensors; ubiquitous computing; pervasive computing

1. Introduction

Since the first Linux-based smartwatch was presented in 2000 at the IEEE International Solid-State Circuits Conference (ISSCC) by Steve Mann, who was later hailed as the "father of wearable computing", the 21st century has witnessed a rapid growth of wearables. For example, as of January 2020, 21% of adults in the United States, most of whom are not opposed to sharing data with medical researchers, own a smartwatch [1].

In addition to being fashion accessories, wearables provide unprecedented opportunities for monitoring human physiological signals and facilitating natural and seamless interaction between humans and machines. Wearables integrate low-power sensors that allow them to sense movement and other physiological signals such as heart rate, temperature, blood pressure, and electrodermal activity. The rapid proliferation of wearable technologies and advancements in sensing analytics have spurred the growth of human activity recognition (HAR). As a general understanding of the HAR shown in Figure 1, HAR has drastically improved the quality of service in a broad range of applications spanning healthcare, entertainment, gaming, industry, and lifestyle, among others. Market analysts

from Meticulous Research® [2] forecast that the global wearable devices market will grow at a compound annual growth rate of 11.3% from 2019, reaching $62.82 billion by 2025, with companies like Fitbit®, Garmin®, and Huawei Technologies® investing more capital into the area.

In the past decade, deep learning (DL) has revolutionized traditional machine learning (ML) and brought about improved performance in many fields, including image recognition, object detection, speech recognition, and natural language processing. DL has improved the performance and robustness of HAR, speeding its adoption and application to a wide range of wearable sensor-based applications. There are two key reasons why DL is effective for many applications. First, DL methods are able to directly learn robust features from raw data for specific applications, whereas features generally need to be manually extracted or engineered in traditional ML approaches, which usually requires expert domain knowledge and a large amount of human effort. Deep neural networks can efficiently learn representative features from raw signals with little domain knowledge. Second, deep neural networks have been shown to be universal function approximators, capable of approximating almost any function given a large enough network and sufficient observations [3–5]. Due to this expressive power, DL has seen a substantial growth in HAR-based applications.

Despite promising results in DL, there are still many challenges and problems to overcome, leaving room for more research opportunities. We present a review on deep learning in HAR with wearable sensors and elaborate on ongoing challenges, obstacles, and future directions in this field.
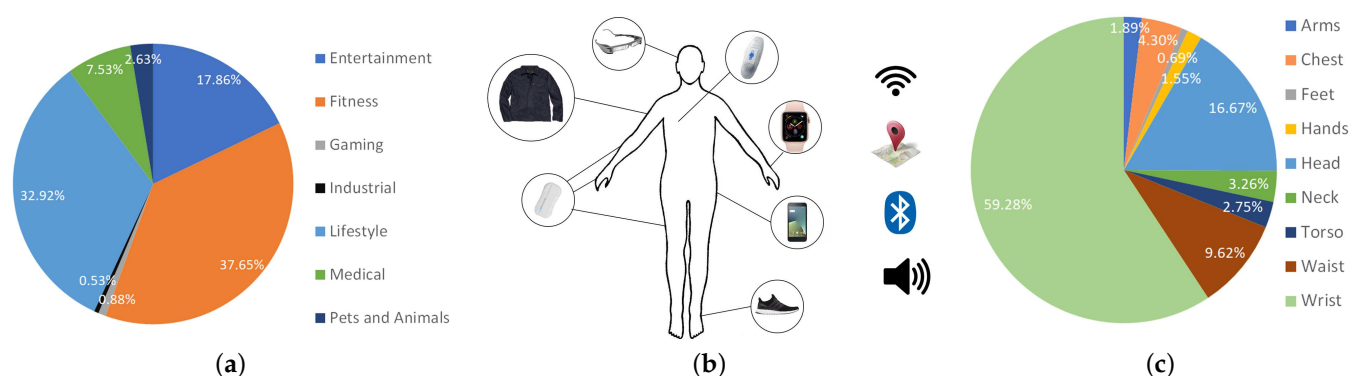

Figure 1. Wearable devices and their application. (**a**) Distribution of wearable applications [6]. (**b**) Typical wearable devices. (**c**) Distribution of wearable devices placed on common body areas [6].

Specifically, we focus on the recognition of physical activities, including locomotion, activities of daily living (ADL), exercise, and factory work. While DL has shown a lot of promise in other applications, such as ambient scene analysis, emotion recognition, or subject identification, we focus on HAR. Throughout this work, we present brief and high-level summaries of major DL methods that have significantly impacted wearable HAR. For more details about specific algorithms or basic DL, we refer the reader to original papers, textbooks, and tutorials [7,8]. Our contributions are summarized as followings.

(i)   Firstly, we give an overview of the background of the human activity recognition research field, including the traditional and novel applications where the research community is focusing, the sensors that are utilized in these applications, as well as widely-used publicly available datasets.

(ii) Then, after briefly introducing the popular mainstream deep learning algorithms, we give a review of the relevant papers over the years using deep learning in human activity recognition using wearables. We categorize the papers in our scope according to the algorithm (autoencoder, CNN, RNN, etc.). In addition, we compare different DL algorithms in terms of the accuracy of the public dataset, pros and cons, deployment, and high-level model selection criteria.

(iii)   We provide a comprehensive systematic review on the current issues, challenges, and opportunities in the HAR domain and the latest advancements towards solutions. At last, honorably and humbly, we make our best to shed light on the possible future directions with the hope to benefit students and young researchers in this field.

2. Methodology

2.1. Research Question

In this work, we propose several major research questions, including Q1 : What the real-world applications of HAR, mainstream sensors, and major public datasets are in this field, Q2 : What deep learning approaches are employed in the field of HAR and what pros and cons each of them have, and Q3 : What challenges we are facing in this field and what opportunities and potential solutions we may have. In this work, we review the state-of-the-art work in this field and present our answers to these questions.

This article is organized as follows: We compare this work with related existing review work in this field in Section 3. Section 4.1 introduces common applications for HAR. Section 4.2 summarizes the types of sensors commonly used in HAR. Section 4.3 summarizes major datasets that are commonly used to build HAR applications. Section 5 introduces the major works in DL that contribute to HAR. Section 6 discusses major challenges, trends, and opportunities for future work. We provide concluding remarks in Section 7.

2.2. Research Scope

In order to provide a comprehensive overview of the whole HAR field, we conducted a systematic review for human activity recognition. To ensure that our work satisfies the requirements of a high-quality systemic review, we conducted the 27-item PRISMA review process [9] and ensured that our work satisfied each requirement. We searched in Google Scholar with meta-keywords (We began compiling papers for this review in November 2020. As we were preparing this review, we compiled a second round of papers in November 2021 to incorporate the latest works published in 2021). (A) "Human activity recognition", "motion recognition", "locomotion recognition", "hand gesture recognition", "wearable", (B) "deep learning", "autoencoder" (alternatively "auto-encoder"), "deep belief network", "convolutional neural network" (alternatively "convolution neural network"), "recurrent neural network", "LSTM", "recurrent neural network", "generative adversarial network" (alternatively "GAN"), "reinforcement learning", "attention", "deep semi-supervised learning", and "graph neural network". We used an AND rule to get combinations of the above meta-keywords (A) and (B). For each combination, we obtained top 200 search results ranked by relevance. We didn't consider any patent or citation-only search result (no content available online).

There are several exclusion criteria to build the database of the paper we reviewed. First of all, we omitted image or video-based HAR works, such as [10], since there is a huge body of work in the computer vision community and the method is significantly different from sensor-based HAR. Secondly, we removed the papers using environmental sensors or systems assisted by environmental sensors such as WiFi- and RFID-based HAR. Thirdly, we removed the papers with minor algorithmic advancements based on prior works. We aim to present the technical progress and algorithmic achievements in HAR, so we avoid presenting works that do not stress the novelty of methods. In the end, as the field of wearable-based HAR is becoming excessively popular and numerous papers are coming out, it is not a surprise to find that many papers share rather similar approaches, and it is almost impossible and less meaningful to cover all of them. Figure 2 shows the consort diagram that outlines step-by-step how we filtered out papers to arrive at the final 176 papers we included in this review. We obtained 8400 papers in the first step by searching keywords mentioned above on Google Scholar. Next, we removed papers that did not align with the topics in this review (i.e., works that do not utilize deep learning in wearable systems), leaving us with 870 papers. In this step, we removed 2194 papers

that utilized vision, 2031 papers that did not use deep learning, and 2173 papers that did not perform human activity recognition. Then, we removed 52 review papers, 109 papers that did not propose novel systems or algorithms, and five papers that were not in English, leaving us with 704 papers. Finally, we selected the top 25% most relevant papers to review, leaving us with 176 papers that we reviewed for this work. We used the relevancy score provided through Google Scholar to select the papers to include in this systemic review. Therefore, we select, categorize, and summarize representative works to present in this review paper. We adhere to the goal of our work throughout the whole paper, that is, to give an overall introduction to new researchers entering this field and present cutting-edge research challenges and opportunities.
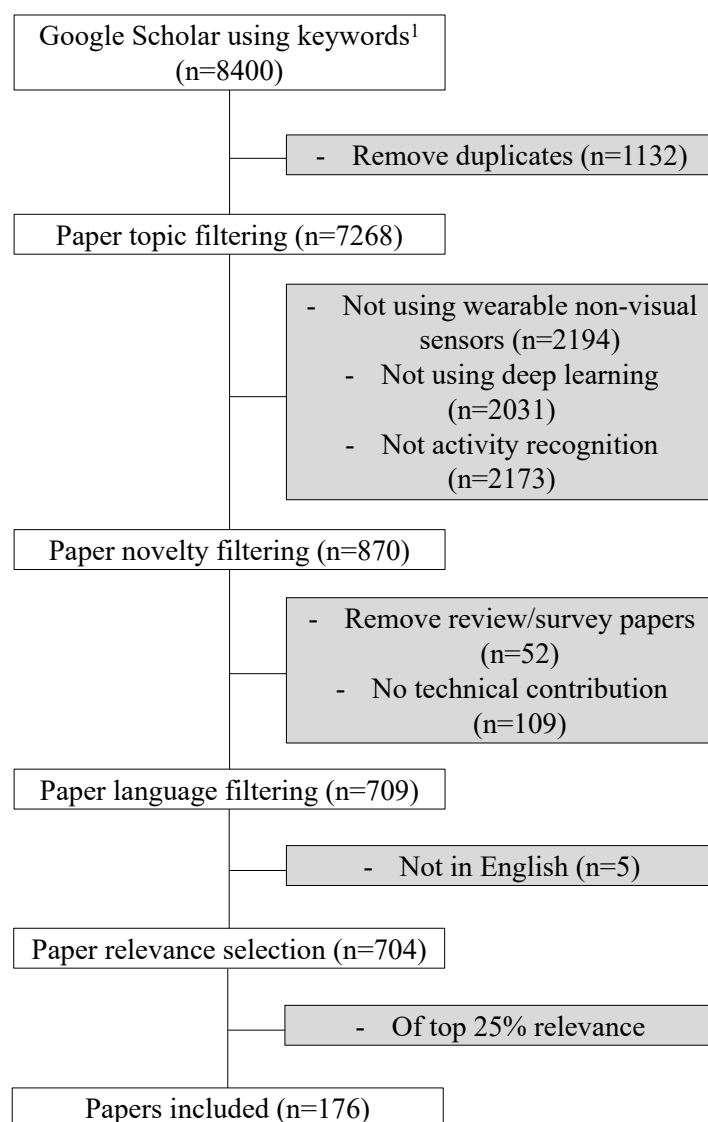
Figure 2. Consort diagram outlining how we selected the final papers we included in this work.

However, we admit that the review process conducted in this work has some limitations. Due to the overwhelming amount of papers in this field in recent years, it is almost impossible to include all the published papers in the field of deep learning-based wearable human activity recognition in a single review paper. The selection of the representative works to present in this paper is unavoidably subject to the risk of bias. Besides, we may miss the very first paper initiating or adopting a certain method. At last, due to the nature of human-related research and machine learning research, many possibilities could

cause heterogeneity among study results, including the heterogeneity in devices, heterogeneity from the demography of participants, and even heterogeneity from the algorithm implementation details.

2.3. Taxonomy of Human Activity Recognition

In order to obtain a straightforward understanding of the hierarchies under the tree of HAR, we illustrate the taxonomy of HAR as shown in Figure 3. We categorized existing HAR works into four dimensions: Sensor, application, DL approach, and challenge. There are basically two kinds of sensors: Physical sensors and physiological sensors. Physical sensors include Inertial Measurement Unit (IMU), piezoelectric sensor, GPS, wearable camera, etc. Some exemplary physiological sensors are electromyography (EMG) and photoplethysmography (PPG), just to name a few. In terms of the applications of HAR systems, we categorized them into healthcare, fitness& lifestyle, and Human Computer Interaction (HCI). Regarding the DL algorithm, we introduce six approaches, including autoencoder (AE), Deep Belief Network (DBN), Convolutional Neural Network, Recurrent Neural Network (including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs)), Generative Adversarial Network (GAN), and Deep Reinforcement Learning (DRL). In the end, we discuss the challenges our research community is facing and the state-of-the-art works are coping with, also shown in Figure 3.
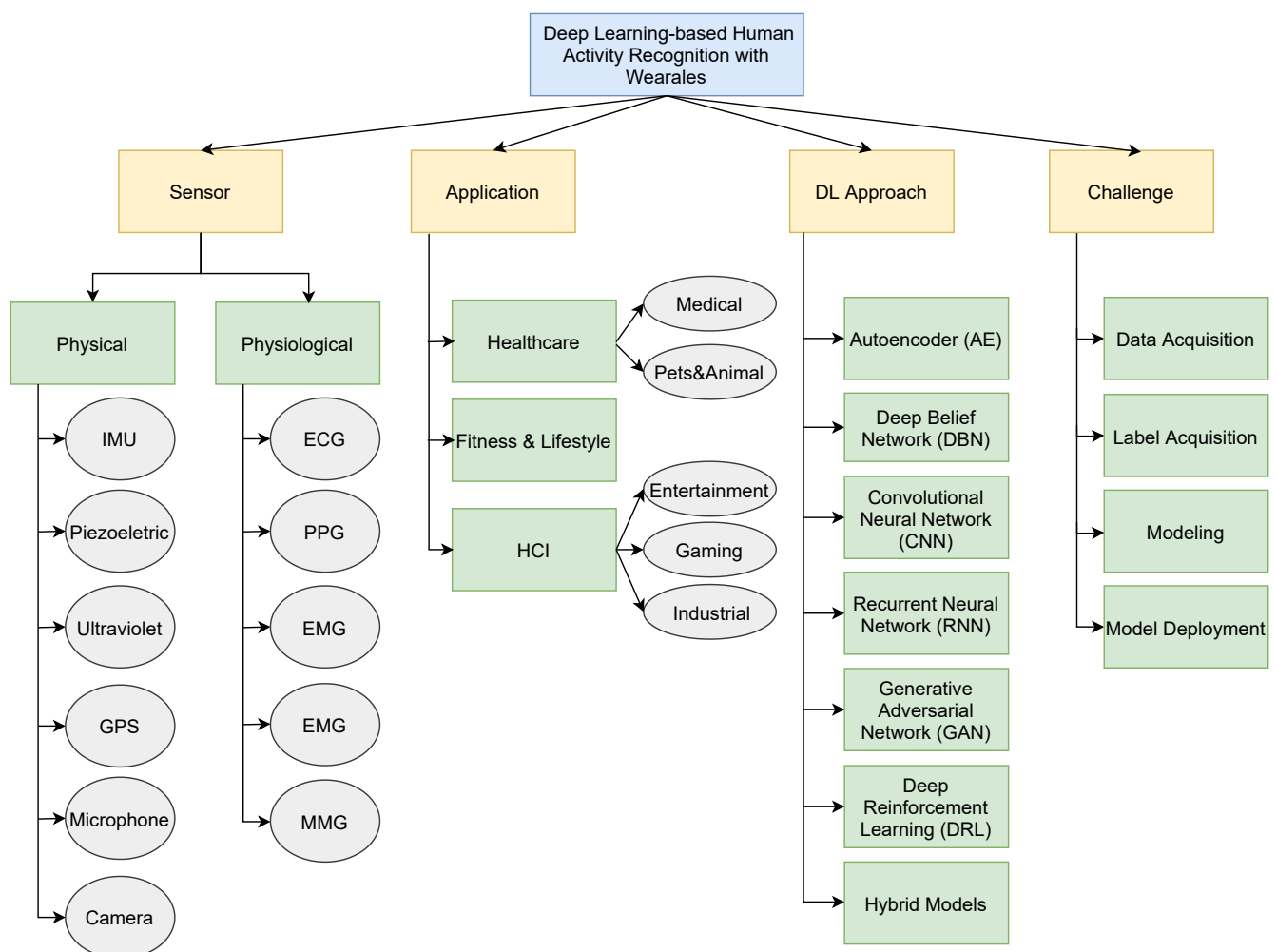

Figure 3. Taxonomy of Deep Learning-based Human Activity Recognition with Wearables.

3. Related Work

There are some existing review papers in the literature for deep learning approaches for sensor-based human activity recognition [11–14]. Nweke et al. accentuated the advancements in deep learning models by proposing a taxonomy of generative, discriminative, and hybrid methods along with further explorations for the advantages and limitations up to year 2018 [12]. Similarly, Wang et al. conducted a thorough analysis on different sensor-based modalities, deep learning models, and their respective applications up to the year 2017 [11]. However, in recent years, due to huge advancements in the availability and computational power of computing resources and cutting-edge deep learning techniques, the applied deep learning area has been revolutionized and reached all-time-high performance in the field of sensor-based human activity recognition. Therefore, we aim to present the most recent advances and most exciting achievements in the community in a timely manner to our readers.

In another work, Chen et al. provided the community a comprehensive review which has done an in-depth analysis of the challenges/opportunities for deep learning in sensor-based HAR and proposed a new taxonomy for the challenges ahead of the activity recognition systems [13]. In contrast, we view our work as more of a gentle introduction of this field to students and novices in the way that our literature review provides the community with a detailed analysis on most recent state-of-the-art deep learning architectures (i.e., CNN, RNN, GAN, Deep Reinforcement Learning, and hybrid models) and their respective pros and cons on HAR benchmark datasets. At the same time, we distill our knowledge and experience from our past works in this field and present the challenges and opportunities from a different viewpoint. Another recent work was presented by Ramanujam et al. , in which they categorized the deep learning architectures in CNN, LSTM, and hybrid methods and conducted an in-depth analysis on the benchmark datasets [14]. Compared with their work, our paper pays more attention to the most recent cutting-edge deep learning methods applied on HAR on-body sensory data, such as GAN and DRL. We also provide both new learners and experienced researchers with a profound resource in terms of model comparison, model selection and model deployment. In a nutshell, our review has thoroughly analysed most up-to-date deep learning architectures applied on various wearable sensors, elaborated on their respective applications, and compared performances on public datasets. What's more, we attempt to cover the most recent advances in resolving the challenges and difficulties and shed light on possible research opportunities.

4. Human Activity Recognition Overview

4.1. Applications

In this section, we illustrate the major areas and applications of wearable devices in HAR. Figure 1a, taken from the wearable technology database [6], breaks down the distribution of application types of 582 commercial wearables registered since 2015 [6]. The database suggests that wearables are increasing in popularity and will impact people's lives in several ways, particularly in applications ranging from fitness and lifestyle to medical and human-computer interaction.

4.1.1. Wearables in Fitness and Lifestyle

Physical activity involves activities such as sitting, walking, laying down, going up or downstairs, jogging, and running [15]. Regular physical activity is increasingly being linked to a reduction in risk for many chronic diseases, such as obesity, diabetes, and cardiovascular disease, and has been shown to improve mental health [16]. The data recorded by wearable devices during these activities include plenty of information, such as duration and intensity of activity, which further reveals an individual's daily habits and health conditions [17]. For example, dedicated products such as Fitbit [18] can estimate and record energy expenditure on smart devices, which can further serve as an important step in tracking personal activity and preventing chronic diseases [19]. Moreover,

there has been evidence of the association between modes of transport (motor vehicle, walking, cycling, and public transport) and obesity-related outcomes [20]. Being aware of daily locomotion and transportation patterns can provide physicians with the necessary information to better understand patients' conditions and also encourage users to engage in more exercise to promote behavior change [21]. Therefore, the use of wearables in fitness and lifestyle has the potential to significantly advance one of the most prolific aspects of HAR applications [22–30].

Energy (or calorie) expenditure (EE) estimation has grown to be an important reason why people care to track their personal activity. Self-reflection and self-regulation of one's own behavior and the habit has been important factor in designing interventions that prevent chronic diseases such as obesity, diabetes, and cardiovascular diseases.

4.1.2. Wearables in Healthcare and Rehabilitation

HAR has greatly impacted the ability to diagnose and capture pertinent information in healthcare and rehabilitation domains. By tracking, storing, and sharing patient data with medical institutions, wearables have become instrumental for physicians in patient health assessment and monitoring. Specifically, several works have introduced systems and methods for monitoring and assessing Parkinson disease (PD) symptoms [31–36]. Pulmonary disease, such as Chronic Obstructive Pulmonary Disease (COPD), asthma, and COVID-19, is one of leading causes of morbidity and mortality. Some recent works use wearables to detect cough activity, a major symptom of pulmonary diseases [37–40]. Other works have introduced methods for monitoring stroke in infants using wearable accelerometers [41] and methods for assessing depressive symptoms utilizing wrist-worn sensors [42]. In addition, detecting muscular activities and hand motions using electromyography (EMG) sensors has been widely applied to enable improved prostheses control for people with missing or damaged limbs [43–46].

4.1.3. Wearables in Human Computer Interaction (HCI)

Modern wearable technology in HCI has provided us with flexible and convenient methods to control and communicate with electronics, computers, and robots. For example, a wrist-worn wearable outfitted with an inertial measurement unit (IMU) can easily detect the wrist shaking [47–49] to control smart devices to skip a song by shaking the hand, instead of bringing up the screen, locating, and pushing a button. Furthermore, wearable devices have played an essential role in many HCI applications in entertainment systems and immersive technology. One example field is augmented reality (AR) and virtual reality (VR), which has changed the way we interact and view the world. Thanks to accurate activity, gesture, and motion detection from wearables, these applications could induce feelings of cold or hot weather by providing an immersive experience by varying the virtual environment and could enable more realistic interaction between the human and virtual objects [43,44].

4.2. Wearable Sensors

Wearable sensors are the foundation of HAR systems. As shown in Figure 1b, there are a large number of off-the-shelf smart devices or prototypes under development today, including smartphones, smartwatches, smart glasses, smart rings [50], smart gloves [51], smart armbands [52], smart necklaces [53–55], smart shoes [56], and E-tattoos [57]. These wearable devices cover the human body from head to toe with a general distribution of devices shown in Figure 1c, as reported by [6]. The advance of micro-electro-mechanical system (MEMS) technology (microscopic devices, comprising a central unit such as a microprocessor and multiple components that interact with the surroundings such as microsensors) has allowed wearables to be miniaturized and lightweight to reduce the burden on adherence to the use of wearables and Internet of Things (IoT) technologies. In this section, we introduce and discuss some of the most prevalent MEMS sensors commonly

used in wearables for HAR. The summary of wearable sensors is represented as a part of Figure 3.

4.2.1. Inertial Measurement Unit (IMU)

Inertial measurement unit (IMU) is an integrated sensor package comprising of accelerometer, gyroscope, and sometimes magnetometer. Specifically, an accelerometer detects linear motion and gravitational forces by measuring the acceleration in 3 axes (x, y, and z), while a gyroscope measures rotation rate (roll, yaw, and pitch). The magnetometer is used to detect and measure the earth's magnetic fields. Since a magnetometer is often used to obtain the posture and orientation in accordance with the geomagnetic field, which is typically outside the scope of HAR, the magnetometer is not always included in data analysis for HAR. By contrast, accelerometers and gyroscopes are commonly used in many HAR applications. We refer to an IMU package comprising a 3-axis accelerometer and a 3-axis gyroscope as a 6-axis IMU. This component is often referred to as a 9-axis IMU if a 3-axis magnetometer is also integrated. Owing to mass manufacturing and the widespread use of smartphones and wearable devices in our daily lives, IMU data are becoming more ubiquitous and more readily available to collect. In many HAR applications, researchers carefully choose the sampling rate of the IMU sensors depending on the activity of interest, often choosing to sample between 10 and several hundred Hz. In [58], Chung et al. tested a range of sampling rates and gave the best one in his application. Besides, it's been shown that higher sampling rates allow the system to capture signals with higher precision and frequencies, leading to more accurate models at the cost of higher energy and resource consumption. For example, the projects presented in [59,60] utilize sampling rates above the typical rate. These works sample at 4 kHz to sense the vibrations generated from the interaction between a hand and a physical object.

4.2.2. Electrocardiography (ECG) and Photoplethysmography (PPG)

Electrocardiography (ECG) and photoplethysmography (PPG) are the most commonly used sensing modalities for heart rate monitoring. ECG, also called EKG, detects the heart's electrical activity through electrodes attached to the body. The standard 12-lead ECG attaches ten non-intrusive electrodes to form 12 leads on the limbs and chest. ECG is primarily employed to detect and diagnose cardiovascular disease and abnormal cardiac rhythms. PPG relies on using a low-intensity infrared (IR) light sensor to measure blood flow caused by the expansion and contraction of heart chambers and blood vessels. Changes in blood flow are detected by the PPG sensor as changes in the intensity of light; filters are then applied to the signal to obtain an estimate of heart rate. Since ECG directly measures the electrical signals that control heart activity, it typically provides more accurate measurements for heart rate and often serves as a baseline for evaluating PPG sensors.

4.2.3. Electromyography (EMG)

Electromyography (EMG) measures the electrical activity produced by muscle movement and contractions. EMG was first introduced in clinical tests to assess and diagnose the functionality of muscles and motor neurons. There are two types of EMG sensors: Surface EMG (sEMG) and intramuscular EMG (iEMG). sEMG uses an array of electrodes placed on the skin to measure the electrical signals generated by muscles through the surface of the skin [61]. There are a number of wearable applications that detect and assess daily activities using sEMG [44,62]. In [63], researchers developed a neural network that distinguishes ten different hand motions using sEMG to advance the effectiveness of prosthetic hands. iEMG places electrodes directly into the muscle beneath the skin. Because of its invasive nature, non-invasive wearable HAR systems do not typically include iEMG.

4.2.4. Mechanomyography (MMG)

Mechanomyography (MMG) uses a microphone or accelerometer to measure low-frequency muscle contractions and vibrations, as opposed to EMG, which uses electrodes.

For example, 4-channel MMG signals from the thigh can be used to detect knee motion patterns [64]. Detecting these knee motions is helpful for the development of power-assisted wearables for powered lower limb prostheses. The authors create a convolutional neural network and support vector machine (CNN-SVM) architecture comprising a seven-layer CNN to learn dominant features for specific knee movements. The authors then replace the fully connected layers with an SVM classifier trained with the extracted feature vectors to improve knee motion pattern recognition. Moreover, Meagher et al. [65] proposed developing an MMG device as a wearable sensor to detect mechanical muscle activity for rehabilitation after stroke.

Other wearable sensors used in HAR include (but are not limited to) piezoelectric sensor [66,67] for converting changes in pressure, acceleration, temperature, strain, or force to electrical charge, barometric pressure sensor [68] for atmospheric pressure, temperature measurement [69], electroencephalography (EEG) for measuring brain activity [70], respiration sensors for breathing monitoring [71], ultraviolet (UV) sensors [72] for sun exposure assessment, GPS for location sensing, microphones for audio recording [39,73,74], and wearable cameras for image or video recording [55]. It is also important to note that the wearable camera market has drastically grown with cameras such as GoPro becoming mainstream [75–78] over the last few years. However, due to privacy concerns posed by participants related to video recording, utilizing wearable cameras for longitudinal activity recognition is not as prevalent as other sensors. Additionally, HAR with image/video processing has been extensively studied in the computer vision community [79,80], and the methodologies commonly used differ significantly from techniques used for IMUs, EEG, PPG, etc. For these reasons, despite their significance in applications of deep learning methods, this work does not cover image and video sensing for HAR.

4.3. Major Datasets

We list the major datasets employed to train and evaluate various ML and DL techniques in Table 1, ranked based on the number of citations they received per year according to Google Scholar. As described in the earlier sections, most datasets are collected via IMU, GPS, or ECG. While most datasets are used to recognize physical activity or daily activities [81–99], there are also a few datasets dedicated to hand gestures [100,101], breathing patterns [102], and car assembly line activities [103], as well as those that monitor gait for patients with PD [104].

Table 1. Major Public Datasets for Wearable-based HAR.

Dataset	Application	Sensor	# Classes	Spl. Rate	Citations/yr
WISDM [81]	Locomotion	3D Acc.	6	20 Hz	217
ActRecTut [100]	Hand gestures	9D IMU	12	32 Hz	153
UCR(UEA)-TSC [105,106]	9 datasets (e.g., uWave [107])	Vary	Vary	Vary	107
UCI-HAR [82]	Locomotion	Smartphone 9D IMU	6	50 Hz	78
Ubicomp 08 [83]	Home activities	Proximity sensors	8	N/A	69
SHO [84]	Locomotion	Smartphone 9D IMU	7	50 Hz	52
UTD-MHAD1/2 [85]	Locomotion & activities	3D Acc. & 3D Gyro.	27	50 Hz	39
HHAR [86]	Locomotion	3D Acc.	6	50–200 Hz	37
Daily & Sports Activities [87]	Locomotion	9D IMU	19	25 Hz	37
MHEALTH [88,89]	Locomotion & gesture	9D IMU & ECG	12	50 Hz	33
Opportunity [90]	Locomotion & gesture	9D IMU	16	50 Hz	32
PAMAP2 [91]	Locomotion & activities	9D IMU & HR monitor	18	100 Hz	32
Daphnet [104]	Freezing of gait	3D Acc.	2	64 Hz	30
SHL [108]	Locomotion & transportation	9D IMU	8	100 Hz	23
SARD [92]	Locomotion	9D IMU & GPS	6	50 Hz	22
Skoda Checkpoint [103]	Assembly-line activities	3D Acc.	11	98 Hz	21

Table 1. *Cont.*

Dataset	Application	Sensor	# Classes	Spl. Rate	Citations/yr
UniMiB SHAR [93]	Locomotion & gesture	9D IMU	12	N/A	20
USC-HAD [94]	Locomotion	3D ACC. & 3D Gyro.	12	100 Hz	20
ExtraSensory [95]	Locomotion & activities	9D IMU & GPS	10	25–40 Hz	13
HASC [96]	Locomotion	Smartphone 9D IMU	6	100 Hz	11
Actitracker [97]	Locomotion	9D IMU & GPS	5	N/A	6
FIC [101]	Feeding gestures	3D Acc.	6	20 Hz	5
WHARF [98]	Locomotion	Smartphone 9D IMU	16	50 Hz	4

Most of the datasets listed above are publicly available. The University of California Riverside-Time Series Classification (UCR-TSC) archive is a collection of datasets collected from various sensing modalities [109]. The UCR-TSC archive was first released and included 16 datasets, growing to 85 datasets by 2015 and 128 by October 2018. Recently, researchers from the University of East Anglia have collaborated with UCR to generate a new collection of datasets, which includes nine categories of HAR: BasicMotions , Cricket, Epilepsy, ERing, Handwriting, Libras, NATOPS, RacketSports, and UWaveGestureLibrary [106]. One of the most commonly used datasets is the OPPORTUNITY dataset [90]. This dataset contains data collected from 12 subjects using 15 wireless and wired networked sensor systems, with 72 sensors and ten modalities attached to the body or the environment. Existing HAR papers mainly focus on data from on-body sensors, including 7 IMUs and 12 additional 3D accelerometers for classifying 18 kinds of activities. Researchers have proposed various algorithms to extract features from sensor signals and to perform activity classification using machine-learned models like K Nearest Neighbor (KNN) and SVM [22,110–118]. Another widely-used dataset is PAMAP2 [91], which is collected from 9 subjects performing 18 different activities, ranging from jumping to house cleaning, with 3 IMUs (100-Hz sampling rate) and a heart rate monitor (9 Hz) attached to each subject. Other datasets such as Skoda [103] and WISDM [81] are also commonly used to train and evaluate HAR algorithms. In Figure 4, we present the placement of inertial sensors in 9 common datasets.
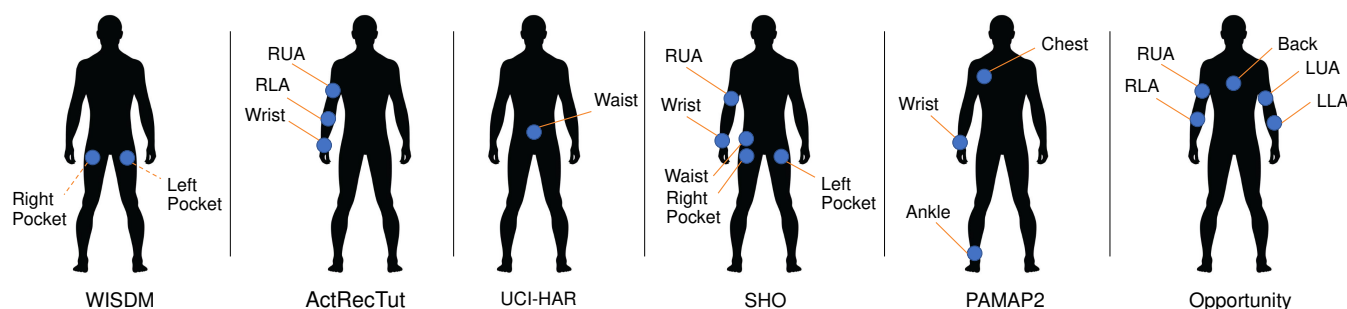

Figure 4. Placement of inertial sensors in different datasets: WISDOM; ActRecTut; UCI-HAR; SHO; PAMAP2; and Opportunity.

5. Deep Learning Approaches

In recent years, DL approaches have outperformed traditional ML approaches in a wide range of HAR tasks. There are three key factors behind deep learning's success: Increasingly available data, hardware acceleration, and algorithmic advancements. The growth of datasets publicly shared through the web has allowed developers and researchers to quickly develop robust and complex models. The development of GPUs and FPGAs have drastically shortened the training time of complex and large models. Finally, improvements in optimization and training techniques have also improved training speed. In this section, we will describe and summarize HAR works from six types of deep learning approaches. We also present an overview of deep learning approaches in Figure 3.

5.1. Autoencoder

The autoencoder, originally called "autoassociative learning module", was first proposed in the 1980s as an unsupervised pre-training method for artificial neural networks (ANN) [119]. Autoencoders have been widely adopted as an unsupervised method for learning features. As such, the outputs of autoencoders are often used as inputs to other networks and algorithms to improve performance [120–124]. An autoencoder is generally composed of an encoder module and a decoder module. The encoding module encodes the input signals into a latent space, while the decoder module transforms signals from the latent space back into the original domain. As shown in Figure 5, the encoder and decoder module is usually several dense layers (i.e., fully connected layers) of the form

$$f_\theta(\mathbf{x}) : \mathbf{z} = \sigma(W_e \mathbf{x} + b_e)$$
$$g_{\theta'}(\mathbf{z}) : \mathbf{x}' = \sigma(W_d \mathbf{z} + b_d)$$

where $\theta = \{W_e, b_e\}$, $\theta' = \{W_d, b_d\}$ are the learnable parameters of the encoder and decoder. σ is the non-linear activation function, such as Sigmoid, tanh, or rectified linear unit (ReLU). W_e and W_d refer to the weights of the layer, while b_e and b_d are the bias vectors. By minimizing a loss function applied on $\mathbf{x}$ and $\mathbf{x}'$, autoencoders aim at generating the final output by imitating the input. Autoencoders are efficient tools for finding optimal codes, $\mathbf{z}$, and performing dimensionality reduction. An autoencoder's strength in dimensionality reduction has been applied to HAR in wearables [34,121,125–131] and functions as a powerful tool for denoising and information retrieval.
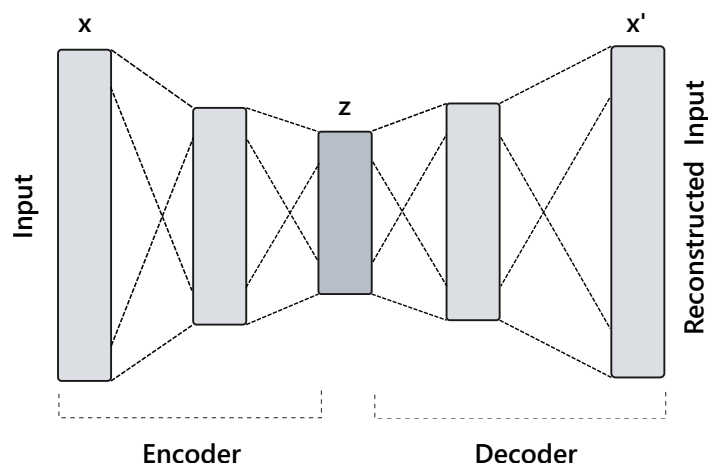

Figure 5. Illustration of an autoencoder network [132].

As such, autoencoders are most commonly used for feature extraction and dimensionality reduction [120,122–126,133–141]. Autoencoders are generally used individually or in a stacked architecture with multiple autoencoders. Mean squared error or mean squared error plus KL divergence loss functions are typically used to train autoencoders. Li et al. presents an autoencoder architecture where a sparse autoencoder and a denoising autoencoder are used to explore useful feature representations from accelerometer and gyroscope sensor data, and then they perform classification using support vector machines [125]. Experiments are performed on a public HAR dataset [82] from the UCI repository, and the classification accuracy is compared with that of Fast Fourier Transform (FFT) in the frequency domain and Principal Component Analysis (PCA). The result reveals that the stacked autoencoder has the highest accuracy of 92.16% and provides a 7% advantage over traditional methods with hand-crafted features. Jun and Choi [142] studied the classification of newborn and infant activities into four classes: Sleeping, moving in agony, moving in normal condition, and movement by an external force. Using the data from an accelerometer attached to the body and a three-layer autoencoder combined with k-means clustering, they achieve 96% weighted accuracy in an unsupervised way. Additionally,

autoencoders have been explored for feature extraction in domain transfer learning [143], detecting unseen data [144], and recognizing null classes [145]. For example, Prabono et al. [146] propose a two-phase autoencoder-based approach of domain adaptation for human activity recognition. In addition, Garcia et al. [147] proposed an effective multiclass algorithm that consists of an ensemble of autoencoders where each autoencoder is associated with a separate class. This modular structure of classifiers makes models more flexible when adding new classes, which only calls for adding new autoencoders instead of re-training the model.

Furthermore, autoencoders are commonly used to sanitize and denoise raw sensor data [127,130,148], a known problem with wearable signals that impacts our ability to learn patterns in the data. Mohammed and Tashev in [127] investigated the use of sensors integrated into common pieces of clothing for HAR. However, they found that sensors attached to loose clothing are prone to contain large amounts of motion artifacts, leading to low mean signal-to-noise ratios (SNR). To remove motion artifacts, the authors propose a deconvolutional sequence-to-sequence autoencoder (DSTSAE). The weights for this network are trained with a weighted form of a standard VAE loss function. Experiments show that the DSTSAE outperforms traditional Kalman Filters and improves the SNR from -12 dB to $+18.2$ dB, with the F1-score of recognizing gestures improved by 14.4% and locomotion activities by 55.3%. Gao et al. explores the use of stacking autoencoders to denoise raw sensor data to improve HAR using the UCI dataset [82,130]. Then, LightGBM (LBG) is used to classify activities using the denoised signals.

Autoencoders are also commonly used to detect abnormal muscle movements, such as Parkinson's Disease and Autism Spectrum Disorder (ASD). Rad et al. in [34] utilizes an autoencoder to denoise and extract optimized features of different movements and use a one-class SVM to detect movement anomalies. To reduce the overfitting of the autoencoder, the authors inject artificial noise to simulate different types of perturbations into the training data. Sigcha et al. in [149] uses a denoising autoencoder to detect freezing of gait (FOG) in Parkinson's disease patients. The autoencoder is only trained using data labelled as a normal movement. During the testing phase, samples with significant statistical differences from training data are classified as abnormal FOG events.

As autoencoders map data into a nonlinear and low-dimensional latent space, they are well-suited for applications requiring privacy preservation. Malekzadeh et al. developed a novel replacement autoencoder that removes prominent features of sensitive activities, such as drinking, smoking, or using the restroom [121]. Specifically, the replacement autoencoder is trained to produce a non-sensitive output from a sensitive input via stochastic replacement while keeping characteristics of other less sensitive activities unchanged. Extensive experiments are performed on Opportunity [90], Skoda [103], and Hand-Gesture [100] datasets. The result shows that the proposed replacement autoencoder can retain the recognition accuracy of non-sensitive tasks using state-of-the-art techniques while simultaneously reducing detection capability for sensitive tasks.

Mohammad et al. introduces a framework called Guardian-Estimator-Neutralizer (GEN) that attempts to recognize activities while preserving gender privacy [128]. The rationale behind GEN is to transform the data into a set of features containing only nonsensitive features. The Guardian, which is constructed by a deep denoising autoencoder, transforms the data into representation in an inference-specific space. The Estimator comprises a multitask convolutional neural network that guides the Guardian by estimating sensitive and non-sensitive information in the transformed data. Due to privacy concerns, it attempts to recognize an activity without disclosing a participant's gender. The Neutralizer is an optimizer that helps the Guardian converge to a near-optimal transformation function. Both the publicly available MobiAct [150] and a new dataset, MotionSense, are used to evaluate the proposed framework's efficacy. Experimental results demonstrate that the proposed framework can maintain the usefulness of the transformed data for activity recognition while reducing the gender classification accuracy to 50% (random guessing) from more than 90% when using raw sensor data. Similarly, the same authors have

proposed another anonymizing autoencoder in [129] for classifying different activities while reducing user identification accuracy. Unlike most works, where the output to the encoder is used as features for classification, this work utilizes both the encoder and decoder outputs. Experiments performed on a self-collected dataset from the accelerometer and gyroscope showcased excellent activity recognition performance (above 92%) while keeping user identification accuracy below 7%.

5.2. Deep Belief Network (DBN)

A DBN, as illustrated in Figure 6, is formed by stacking multiple simple unsupervised networks, where the hidden layer of the preceding network serves as the visible layer for the next. The representation of each sub-network is generally the restricted Boltzmann machine (RBM), an undirected generative energy-based model with a "visible" input layer, a hidden layer, and intra-layer connections in between. The DBN typically has connections between the layers but not between units within each layer. This structure leads to a fast and layer-wise unsupervised training procedure, where contrastive divergence (a training technique to approximate the relationship between a network's weights and its error) is applied to every pair of layers in the DBN architecture sequentially, starting from the "lowest" pair.
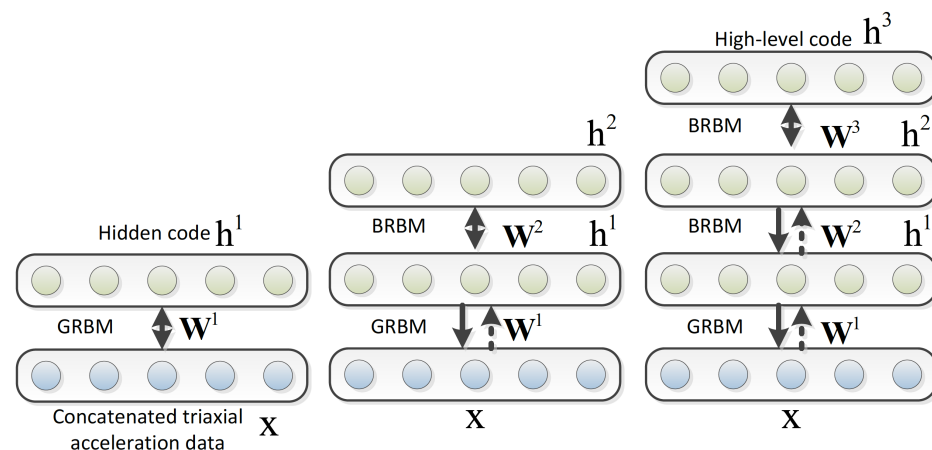

Figure 6. The greedy layer-wise training of DBNs. The first level is trained on triaxial acceleration data. Then, more RBMs are repeatedly stacked to form a deep activity recognition model [151].

The observation that DBNs can be trained greedily led to one of the first effective deep learning algorithms [152]. There are many attractive implementations and uses of DBNs in real-life applications such as drug discovery [153], natural language understanding [154], fault diagnosis [155], etc. There are also many attempts to perform HAR with DBNs. In early exploratory work back in 2011 [156], a five-layer DBN is trained with the input acceleration data collected from mobile phones. The accuracy improvement ranges from 1% to 32% when compared to traditional ML methods with manually extracted features.

In later works, DBN is applied to publicly available datasets [151,157–159]. In [157], two five-layer DBNs with different structures are applied to the Opportunity dataset [90], USC-HAD dataset [94], and DSA dataset [87], and the results demonstrate improved accuracy for HAR over traditional ML methods for all the three datasets. Specifically, the accuracy for the Opportunity, USC-HAD, and DSA datasets are 82.3% (1.6% improvement over traditional methods), 99.2% (13.9% improvement), and 99.1% (15.9% improvement), respectively. In addition, Alsheikh et al. [151] tested the activity recognition performance of DBNs using different parameter settings. Instead of using the raw acceleration data similar to [156], they used spectrogram signals of the triaxial accelerometer data to train the deep activity recognition models. They found that deep models with more layers outperform the shallow models, and the topology of layers having more neurons than the input layer is shown to be more advantageous, which indicates overcompete representation

is essential for learning deep models. The accuracy of the tuned DBN was 98.23%, 91.5%, and 89.38% on the WISDM [81], Daphnet [104], and Skoda [103] benchmark datasets, respectively. In [158], a RBM is used to improve upon other methods of sensor fusion, as neural networks can identify non-intuitive predictive features largely from cross-sensor correlations and thus offer a more accurate estimation. The recognition accuracy with this architecture on the Skoda dataset reached 81%, which is around 6% higher than the traditional classification method with the best performance (Random Forest).

In addition to taking advantage of public datasets, there are also researchers employing DBNs on human activity or health-related recognition with self-collected datasets [31,160]. In [31], DBNs are employed in Parkinson's disease diagnosis to explore if they can cope with the unreliable labelling that results from naturalistic recording environments. The data was collected with two tri-axial accelerometers, with one worn on each wrist of the participant. The DBNs built are two-layer RBMs, with the first layer as a Guassian-binary RBM (containing gaussian visible units) and the second layer as binary-binary (containing only binary units) (please refer to [161] for details). In [160], an unsupervised five-layer DBM-DNN is applied for the automatic detection of eating episodes via commercial blue-tooth headsets collecting raw audio signals, and demonstrate classification improvement even in the presence of ambient noise. The accuracy of the proposed DBM-DNN approach is 94%, which is significantly better than SVM with a 75.6% accuracy.

5.3. Convolutional Neural Network (CNN)

A CNN comprises convolutional layers that make use of the convolution operation, pooling layers, fully connected layers, and an output layer (usually Softmax layer). The convolution operation with a shared kernel enables the learning process of space invariant features. Because each filter in a convolutional layer has a defined receptive field, CNN is good at capturing local dependency, compared with a fully-connected neural network. Though each kernel in a layer covers a limited size of input neurons, by stacking multiple layers, the neurons of higher layers will cover a larger more global receptive field. The pyramid structure of CNN contributes to its capability of gathering low-level local features into high-level semantic meanings. This allows CNN to learn excellent features as shown in [162], which compares the features extracted from CNN to hand-crafted time and frequency domain features (Fast Fourier Transform and Discrete Cosine Transform).

CNN incorporates a pooling layer that follows each convolutional layer in most cases. A pooling layer compresses the representation it is learning and strengthens the model against noise by dropping a portion of the output to a convolutional layer. Generally, a few fully connected layers follow after a stack of convolutional and pooling layers that reduce feature dimensionality before being fed into the output layer. A softmax classifier is usually selected as the final output layer. However, as an exception, some studies explored the use of traditional classifiers as the output layer in a CNN [64,118].

Most CNNs use univariate or multivariate sensor data as input. Besides raw or filtered sensor data, the magnitude of 3-axis acceleration is often used as input, as shown in [26]. Researchers have tried encoding time-series data into 2D images as input into the CNN. In [163], the Short-time Fourier transform (STFT) for time-series sensor data is calculated, and its power spectrum is used as the input to a CNN. Since time series data is generally one-dimensional, most CNNs adopt 1D-CNN kernels. Works that use frequency-domain inputs (e.g., spectrogram), which have an additional frequency dimension, will generally use 2D-CNN kernels [164]. The choice of 1D-CNN kernel size normally falls in the range of 1×3 to 1×5 (with exceptions in [22,63,64] where kernels of size 1×8, 2×101, and 1×20 are adopted).

To discover the relationship between the number of layers, the kernel size, and the complexity level of the tasks, we picked and summarized several typical studies in Table 2. A majority of the CNNs consist of five to nine layers [23,63,64,113,114,165–168], usually including two to three convolutional layers, two to three max-pooling layers, followed by one to two fully connected layers before feeding the feature representation into the

output layer (softmax layer in most cases). Dong et al. [169] demonstrated performance improvements by leveraging both handcrafted time and frequency domain features along with features generated from a CNN, called HAR-Net, to classify six locomotion activities using accelerometer and gyroscope signals from a smartphone. Ravi et al. [170] used a shallow three-layer CNN network including a convolutional layer, a fully connected layer, and a softmax layer to perform on-device activity recognition on a resource-limited platform and shown its effectiveness and efficiency on public datasets. Zeng et al. [22] and Lee et al. [26] also used a small number of layers (four layers). The choice of the loss function is an important decision in training CNNs. In classification tasks, cross-entropy is most commonly used, while in regression tasks, mean squared error is most commonly used. Most CNN models process input data by extracting and learning channel-wise features separately while Huang et al. [167] first propose a shallow CNN that considers cross-channel communication. The channels in the same layer interact with each other to obtain discriminative features of sensor data.

Table 2. Summary of typical studies that use layer-by-layer CNN structure in HAR and their configurations. We aim to present the relationship of CNN kernels, layers, and targeted problems (application and sensors). Key: C—convolutional layer; P—max-pooling layer; FC—fully connected layer; S—softmax; S1—accelerometer; S2—gyroscope; S3—magnetometer; S4—EMG; S5—ECG

Study	Architecture	Kernel Conv.	Application	# Classes	Sensors	Dataset
[26]	C-P-FC-S	$1 \times 3, 1 \times 4, 1 \times 5$	locomotion activities	3	S1	Self
[171]	C-P-C-P-S	4×4	locomotion activities	6, 12	S1	UCI, mHealth
[22]	C-P-FC-FC-S	1×20	daily activities, locomotion activities	-	-	Skoda, Opportunity, Actitracker
[172]	C-P-C-P-FC-S	5×5	locomotion activities	6	S1	WISDM
[173]	C-P-C-P-C-FC	$1 \times 5, 1 \times 9$	locomotion activities	12	S5	mHealth
[174]	C-P-C-P-FC-FC-S	-	daily activities, locomotion activities	12	S1, S2, S3 ECG	mHealth
[175]	C-P-C-P-C-P-S	12×2	daily activities including brush teeth, comb hair, get up from bed, etc	12	S1, S2, S3	WHARF
[23]	C-P-C-P-C-P-S	12×2	locomotion activities	8	S1	Self
[113]	C-P-C-P-U-FC-S, U: unification layer	$1 \times 3, 1 \times 5$	daily activities, hand gesture	18 (Opp) 12 (hand)	S1, S2 (1 for each)	Opportunity Hand Gesture
[63]	C-C-P-C-C-P-FC	1×8	hand motion classification	10	S4	Rami EMG Dataset
[114]	C-C-P-C-C-P-FC-FC-S (one branch for each sensor)	1×5	daily activities, locomotion activities, industrial ordering picking recognition task	18 (Opp) 12 (PAMAP2)	S1, S2, S3	Opportunity, PAMAP2, Order Picking
[163]	C-P-C-P-C-P-FC-FC-FC-S	$1 \times 4, 1 \times 10, 1 \times 15$	locomotion activities	6	S1, S2, S3	Self

The number of sensors used in a HAR study can vary from a single one to as many as 23 [90]. In [23], a single accelerometer is used to collect data from three locations on the body: Cloth pocket, trouser pocket and waist. The authors collect data on 100 subjects, including eight activities such as falling, running, jumping, walking, walking quickly,

step walking, walking upstairs, and walking downstairs. Moreover, HAR applications can involve multiple sensors of different types. To account for all these different types of sensors and activities, Grzeszick et al. [176] proposed a multi-branch CNN architecture. A multi-branch design adopts a parallel structure that trains separate kernels for each IMU sensor and concatenates the output of branches at a late stage, after which one or more fully connected layers are applied on the flattened feature representation before feeding into the final output layer. For instance, a CNN-IMU architecture contains m parallel branches, one per IMU. Each branch contains seven layers, then the outputs of each branch are concatenated and fed into a fully connected and a softmax output layer. Gao et al. [177] has introduced a novel dual attention module including channel and temporal attention to improving the representation learning ability of a CNN model. Their method has outperformed regular CNN considerably on a number of public datasets such as PAMAP2 [91], WISDM [81], UNIMIB SHAR [93], and Opportunity [90].

Another advantage of DL is that the features learned in one domain can be easily generalized or transferred to other domains. The same human activities performed by different individuals can have drastically different sensor readings. To address this challenge, Matsui et al. [163] adapted their activity recognition to each individual by adding a few hidden layers and customizing the weights using a small amount of individual data. They were able to show a 3% improvement in recognition performance.

5.4. Recurrent Neural Network (RNN)

Initially, the idea of using temporal information was proposed in 1991 [178] to recognize a finger alphabet consisting of 42 symbols and in 1995 [179] to classify 66 different hand shapes with about 98% accuracy. Since then, the recurrent neural network (RNN) with time series as input has been widely applied to classify human activities or estimate hand gestures [180–187].

Unlike feed-forward neural networks, an RNN processes the input data in a recurrent behavior. Equivalent to a directed graph, RNN exhibits dynamic behaviors and possesses the capability of modelling temporal and sequential relationships due to a hidden layer with recurrent connections. A typical structure for an RNN is shown in Figure 7 with the current input, x_t, and previous hidden state, h_{t-1}. The network generates the current hidden state, h_t, and output, y_t, is as follows:

$$
\begin{aligned}
h_t &= \mathscr{F}(W_h h_{t-1} + U_h x_t + b_h) \\
y_t &= \mathscr{F}(W_y h_t + b_y)
\end{aligned}
\tag{1}
$$

where W_h, U_h, and W_y are the weights for the hidden-to-hidden recurrent connection, input-to-hidden connection, and hidden-to-output connection, respectively. b_h and b_y are bias terms for the hidden and output states, respectively. Furthermore, each node is associated with an element-wise non-linearity function as an activation function $\mathscr{F}$ such as the sigmoid, hyperbolic tangent (tanh), or rectified linear unit (ReLU).

In addition, many researchers have undertaken extensive work to improve the performance of RNN models in the context of human activity recognition and have proposed various models based on RNNs, including Independently RNN (IndRNN) [188], Continuous Time RNN (CTRNN) [189], Personalized RNN (PerRNN) [190], Colliding Bodies Optimization RNN (CBO-RNN) [191]. Unlike previous models with one-dimension time-series input, Lv et al. [192] builds a CNN + RNN model with stacked multisensor data in each channel for fusion before feeding into the CNN layer. Ketykó et al. [193] uses an RNN to address the domain adaptation problem caused by intra-session, sensor placement, and intra-subject variances.

HAR improves with longer context information and longer temporal intervals. However, this may result in vanishing or exploding gradient problems while backpropagating gradients [194]. In an effort to address these challenges, long short-term memory (LSTM)-based RNNs [195], and Gated Recurrent Units (GRUs) [196] are introduced to model temporal sequences and their broad dependencies. The GRU introduces a reset and update

gate to control the flow of inputs to a cell [197–201]. The LSTM has been shown capable of memorizing and modelling the long-term dependency in data. Therefore, LSTMs have taken a dominant role in time-series and textual data analysis. It has made substantial contributions to human activity recognition, speech recognition, handwriting recognition, natural language processing, video analysis, etc. As illustrated in Figure 7 [202], a LSTM cell is composed of: (1) input gate, i_t, for controlling flow of new information; (2) forget gate, f_t, setting whether to forget content according to internal state; (3) output gate, o_t, controlling output information flow; (4) input modulation gate, g_t, as main input; (5) internal state, c_t, dictates cell internal recurrence; (6) hidden state, h_t, contains information from samples encountered within the context window previously. The relationship between these variables are listed as Equation (2) [202].

$$\begin{cases} i_t = \sigma(b_i + U_i x_t + W_i h_{t-1}) \\ f_t = \sigma\left(b_f + U_f x_t + W_f x_{t-1}\right) \\ o_t = \sigma(b_o + U_o x_t + W_o h_{t-1}) \\ g_t = \sigma\left(b_g + U_g x_t + W_g h_{t-1}\right) \\ c_t = f_t c_{t-1} + g_t i_t \\ h_t = \tanh(c_t) o_t \end{cases} \tag{2}$$
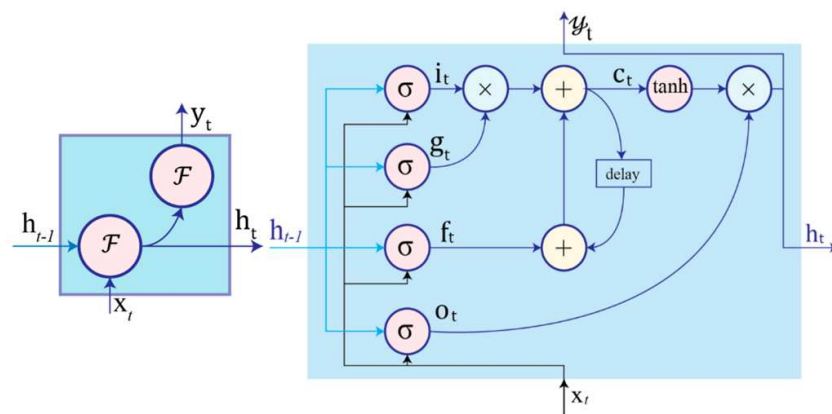

Figure 7. Schematic diagram of an RNN node and LSTM cell [202]. Left: RNN node where h_{t-1} is the previous hidden state, x_t is the current input sample data, h_t is the current hidden state, y_t is the current output, and $\mathcal{F}$ is the activation function. Right: LSTM cell with internal recurrence c_t and outer recurrence h_t.

As shown in Figure 8, the input time series data is segmented into windows and fed into the LSTM model. For each time step, the model computes class prediction scores, which are then merged via late-fusion and used to calculate class membership probabilities through the softmax layer. Previous studies have shown that LSTMs have high performance in wearable HAR [199,202,203]. Researchers in [204] rigorously examine the impact of hyperparameters in LSTM with the fANOVA framework across three representative datasets, containing movement data captured by wearable sensors. The authors assessed thousands of settings with random hyperparameters and provided guidelines for practitioners seeking to apply deep learning to their own problem scenarios [204]. Bidirectional LSTMs, having both past and future recurrent connections, were used in [205,206] to classify activities.
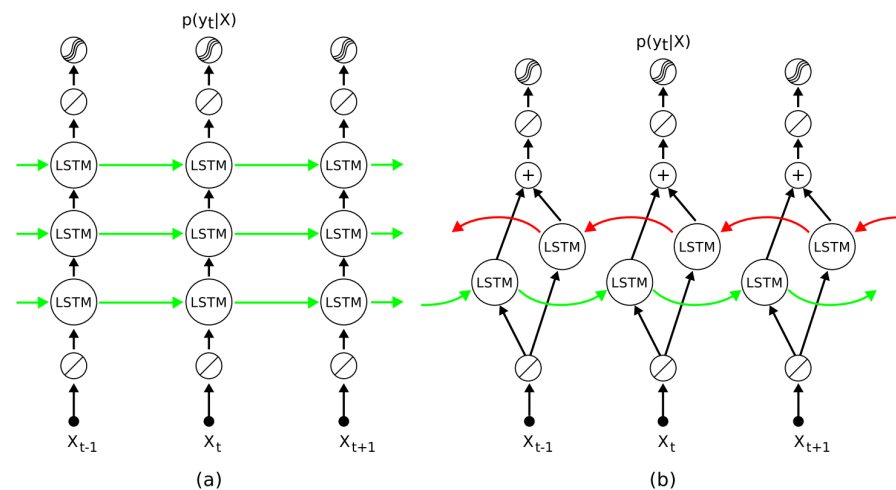
Figure 8. The structure of LSTM and bi-directional LSTM model [204]. (**a**). LSTM network hidden layers containing LSTM cells and a final softmax layer at the top. (**b**) bi-directional LSTM network with two parallel tracks in both future (green) and past (red) directions.

Researchers have also explored other architectures involving LSTMs to improve benchmarks on HAR datasets. Residual networks possess the advantage that they are much easier to train as the addition operator enables gradients to pass through more directly. Residual connections do not impede gradients and could help to refine the output of layers. For example, [200] proposes a harmonic loss function and [207] combines LSTM with batch normalization to achieve 92% accuracy with raw accelerometer and gyroscope data. Ref. [208] proposes a hybrid CNN and LSTM model (DeepConvLSTM) for activity recognition using multimodal wearable sensor data. DeepConvLSTM performed significantly better in distinguishing closely-related activities, such as "Open/Close Door" and "Open/Close Drawer". Moreover, Multitask LSTM is developed in [209] to first extract features with shared weight, and then classify activities and estimate intensity in separate branches. Qin et al. proposed a deep-learning algorithm that combines CNN and LSTM networks [210]. They achieved 98.1% accuracy on the SHL transportation mode classification dataset with CNN-extracted and hand-crafted features as input. Similarly, other researchers [211–219] have also developed the CNN-LSTM model in various application scenarios by taking advantage of the feature extraction ability of CNN and the time-series data reasoning ability of LSTM. Interestingly, utilizing CNN and LSTM combined model, researchers in [219] attempt to eliminate sampling rate variability, missing data, and misaligned data timestamps with data augmentation when using multiple on-body sensors. Researchers in [220] explored the placement effect of motion sensors and discovered that the chest position is ideal for physical activity identification.

Raw IMU and EMG time series data are commonly used as inputs to RNNs [193,221–225]. A number of major datasets used to train and evaluate RNN models have been created, including the Sussex-Huawei Locomotion-Transportation (SHL) [188,198], PAMAP2 [192,226] and Opporunity [203]. In addition to raw time series data [199], Besides raw time series data, custom features are also commonly used as inputs to RNNs. Ref. [197] showed that training an RNN with raw data and with simple custom features yielded similar performance for gesture recognition (96.89% vs 93.38%).

However, long time series may have many sources of noise and irrelevant information. The concept of attention mechanism was proposed in the domain of neural machine translation to address the problem of RNNs being unable to remember long-term relationships. The attention module mimics human visual attention to building direct mappings between the words/phrases that represent the same meaning in two languages. It eliminates the interference from unrelated parts of the input when predicting the output. This is similar to what we as humans perform when we translate a sentence or see a picture for the first time; we tend to focus on the most prominent and central parts of the picture. An RNN

encoder attention module is centred around a vector of importance weights. The weight vector is computed with a trainable feedforward network and is combined with RNN outputs at all the time steps through the dot product. The feedforward network takes all the RNN immediate outputs as input to learn the weights for each time step. [201] utilizes attention in combination with a 1D CNN Gated Recurrent Units (GRUs), achieving HAR performances of 96.5% ± 1.0%, 93.1% ± 2.2%, and 89.3% ± 1.3% on Heterogeneous [86], Skoda [103], and PAMAP2 [91] datasets, respectively. [226] applies temporal attention and sensor attention into LSTM to improve the overall activity recognition accuracy by adaptively focusing on important time windows and sensor modalities.

In recent years, block-based modularized DL networks have been gaining traction. Some examples are GoogLeNet with an Inception module and Resnet with residual blocks. The HAR community is also actively exploring the application of block-based networks. In [227], the authors have used GoogLeNet's Inception module combined with a GRU layer to build a HAR model. The proposed model was showed performance improvements on three public datasets (Opportunity, PAMAP2 and Smartphones datasets). Qian et al. [228] developed the model with SMM_{AR} in a statistical module to learn all orders of moments statistics as features, LSTM in a spatial module to learn correlations among sensors placements, and LSTM + CNN in a temporal module to learn temporal sequence dependencies along the time scale.

5.5. Deep Reinforcement Learning (DRL)

AE, DBN, CNN, and RNN fall within the realm of supervised or unsupervised learning. Reinforcement learning is another paradigm where an agent attempts to learn optimal policies for making decisions in an environment. At each time step, the agent takes an action and then receives a reward from the environment. The state of the environment accordingly changes with the action made by the agent. The goal of the agent is to learn the (near) optimal policy (or probability of action, state pairs) through the interaction with the environment in order to maximize a cumulative long-term reward. The two entities—agent and environment—and the three key elements—action, state and reward—collectively form the paradigm of RL. The structure of RL is shown in Figure 9.
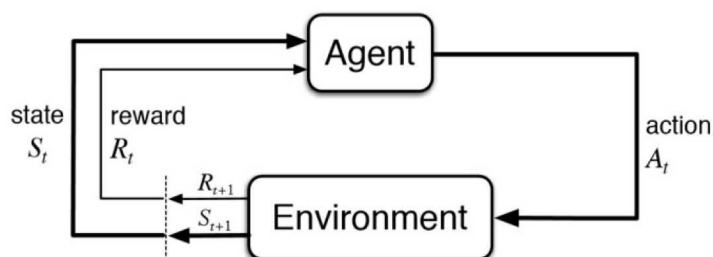

Figure 9. A typical structure of a reinforcement learning network [229].

In the domain of HAR, [230] uses DRL to predict arm movements with 98.33% accuracy. Ref. [231] developed a reinforcement learning model for imitating the walking pattern of a lower-limb amputee on a musculoskeletal model. The system showed 98.02% locomotion mode recognition accuracy. Having a high locomotion recognition accuracy is critical because it helps lower-limb amputees prevent secondary impairments during rehabilitation. In [232], Bhat et al. propose a HAR online learning framework that takes advantage of reinforcement learning utilizing a policy gradient algorithm for faster convergence achieving 97.7% in recognizing six activities.

5.6. Generative Adversarial Network (GAN)

Originally proposed to generate credible fake images that resemble the images in the training set, GAN is a type of deep generative model, which is able to create new samples after learning from real data [233]. It comprises two networks, the generator (**G**) and the discriminator (**D**), competing against each other in a zero-sum game framework as shown

in Figure 10. During the training phase, the generator takes as input a random vector $\mathbf{z}$ and transforms $\mathbf{z} \in \mathbb{R}^{\mathbf{n}}$ to plausible synthetic samples $\hat{x}$ to challenge the discriminator to differentiate between original samples $\mathbf{x}$ and fake samples $\hat{x}$. In this process, the generator strives to make the output probability $\mathbf{D}(\mathbf{G}(\mathbf{z}))$ approach one, in contrast with the discriminator, which tries to make the function's output probability as close to zero as possible. The two adversarial rivals are optimized by finding the Nash equilibrium of the game in a zero-sum game setting, which means the adversarial rivals' gains would be maintained regardless of what strategies are selected. However, it is not theoretically guaranteed that GAN zero-sum games reach Nash Equilibria [234].
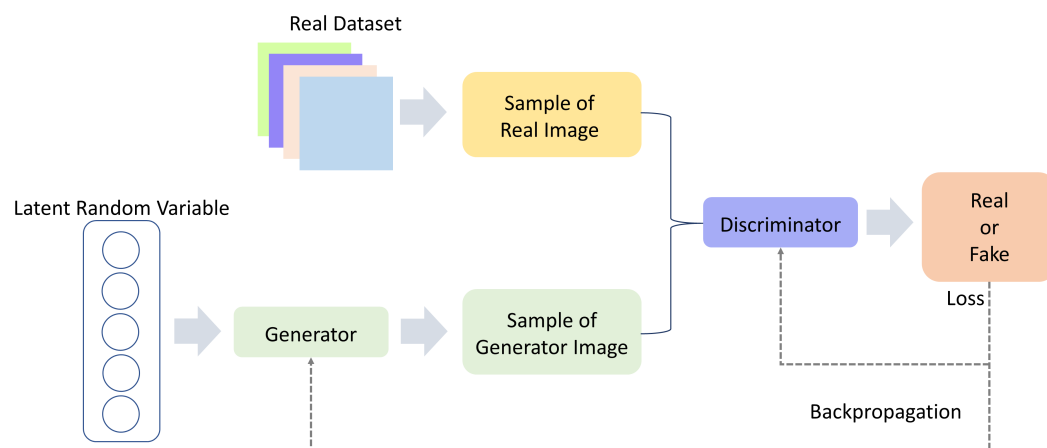

Figure 10. The structure of generative adversarial network.

GAN model has shown remarkable performance in generating synthetic data with high quality and rich details [235,236]. In the field of HAR, GAN has been applied as a semi-supervised learning approach to deal with unlabeled or partially labelled data for improving performance by learning representations from the unlabeled data, which later will be utilized by the network to generalize to the unseen data distribution [237]. Afterwards, GAN has shown the ability to generate balanced and realistic synthetic sensor data. Wang et al. [238] utilized GANs with a customized network to generate synthetic data from the public HAR dataset HASC2010corpus [239]. Similarly, Alharbi et al. [240] assessed synthetic data with CNN or LSTM models as a generator. In two public datasets, Sussex-Huawei Locomotion (SHL) and Smoking Activity Dataset (SAD), the discriminator was built with CNN layers, and the results demonstrated synthetic data with high quality and diversity with two public datasets. Moreover, by oversampling and adding synthetic sensor data into the training, researchers augmented and alleviated the originally imbalanced training set to achieve better performance. In [241,242], they generated verisimilar data of different activities, and Shi et al. [243] used the Boulic kinematic model, which aims to capture the three-dimensional positioning trend to synthesize personified walking data. Due to the ability to generate new data, GAN has been widely applied in transfer learning in HAR to help with the dramatic performance drop when the pre-trained model are tested against unseen data from new users. In transfer learning techniques, the learned knowledge from the source domain (subject) is transferred to the target domain to decrease the lack of performance of the models within the target domain. Moreover, [244] is an attempt that utilized GAN to perform cross-subject transfer learning for HAR since collecting data for each new user was infeasible. With the same idea, cross-subject transfer learning based on GAN outperformed those without GAN on Opportunity benchmark dataset in [244] and outperformed unsupervised learning on UCI and USC-HAD dataset [245]. Even more, transfer learning under conditions of cross-body, cross-user, and cross-sensor has been demonstrated superior performance in [246].

However, much more effort is needed in generating verisimilar data to alleviate the burden and cost of collecting sufficient user data. Additionally, it is typically challenging to

obtain well-trained GAN models owing to the wide variability in amplitude, frequency, and period of the signals obtained from different types of activities.

5.7. Hybrid Models

As an advancement of machine learning models, researchers take advantage of different methods and propose hybrid models. The combination of CNN and LSTM endows the model capability of extracting local features as well as long-term dependencies in sequential data, especially for HAR time series data. For example, Challa et al. [247] proposed a hybrid of CNN and bidirectional long short-term memory (BiLSTM). The accuracy on UCI-HAR, WISDM [81], and PAMAP2 [91] datasets achieved 96.37%, 96.05%, and 94.29%, respectively. Dua et al. [248] proposed a model with CNN combined with GRU and obtained an accuracy of 96.20%, 97.21%, and 95.27% on UCI-HAR, WISDM [81], and PAMAP2 [91] datasets, respectively. In order to have a straightforward view of the functionality of hybrid models, we list several papers with CNN only, LSTM only, CNN + GRU, and CNN + LSTM in Tables 3 and 4. In addition, Zhang et al. [249] proposed to combine reinforcement learning and LSTM model to improve the adaptability of different kinds of sensors, including EEG (EID dataset), RFID (RSSI dataset) [250], and wearable IMU (PAMAP2 dataset) [91]. Ref. [251] employed CNN for feature extraction and a reinforced selective attention model to automatically choose the most characteristic information from multiple channels.

Table 3. Comparison of models on UCI-HAR dataset.

Model	F1-Score (%)	Accuracy (%)
CNN [252]	92.93	92.71
Res-LSTM [253]	91.50	91.60
Stacked-LSTM [254]	–	93.13
CNN-LSTM [215]	–	92.13
Bidir-LSTM [255]	–	92.67
Residual-BiLSTM [253]	93.5	93.6
LSTM-CNN [211]	–	95.78
CNN-GRU [248]	–	96.20
CNN-GRU [247]	94.54	94.58
CNN-LSTM [247]	94.76	94.80
CNN-BiLSTM [247]	96.31	96.37

Table 4. Comparison of models on PAMAP2 dataset.

Model	F1-Score (%)	Accuracy (%)
CNN[252]	91.16	91.00
BiLSTM [255]	89.40	89.52
LSTM-F [204]	92.90	–
COND-CNN [256]	–	94.01
CNN-GRU [248]	–	95.27
CNN-GRU [247]	93.16	93.20
CNN-LSTM [247]	92.77	92.81
CNN-BiLSTM [247]	94.27	94.29

5.8. Summary and Selection of Suitable Methods

Since the last decade, DL methods have gradually dominated a number of artificial intelligence areas, including sensor-based human activity recognition, due to its automatic feature extraction capability, strong expressive power, and the high performance rendered. When a sufficient amount of data are available, we are becoming prone to turn to DL methods. With all these types of available DL approaches discussed above, we need to get a full understanding of the pros and cons of these approaches in order to select

the appropriate approach wisely. To this end, we briefly analyze the characteristics of each approach and attempt to give readers high-level guidance on how to choose the DL approach according to the needs and requirements.

The most salient characteristic of auto-encoder is that it does not require any annotation. Therefore, it is widely adopted in the paradigm of unsupervised learning. Due to its exceptional capability in dimension reduction and noise suppression, it is often leveraged to extract low-dimensional feature representation from raw input. However, auto-encoders may not necessarily learn the correct and relevant characteristics of the problem at hand. There is also generally little insight that can be gained for sensor-based auto-encoders, making it difficult to know which parameters to adjust during training. Deep belief networks are a generative model generally used for solving unsupervised tasks by learning low-dimensional features. Today, DBNs have been less often chosen compared with other DL approaches and are rarely used due to the tedious training process and increased training difficulty with DBN when the network goes deeper [7].

CNN architecture is powerful to extract hierarchical features owing to its layer-by-layer hierarchical structure. When compared with other approaches like RNN and GAN, CNN is relatively easy to implement. Besides, as one of the most studied DL approaches in image processing and computer vision, there is a large range of CNN variants existing that we can choose from to transfer to sensor-based HAR applications. When sensor data are represented as two-dimensional input, we can directly start with pre-trained models on a large image dataset (e.g., ImageNet) to fasten the convergence process and achieve better performance. Therefore, adapting the CNN approach enjoys a higher degree of flexibility in the available network architecture (e.g., GoogLeNet, MobileNet, ResNet, etc) than other DL approaches. However, CNN architecture has the requirement of fixed-sized input, in contrast to RNN, which accepts flexible input size. In addition, compared with unsupervised learning methods such as auto-encoder and DBN, a large number of annotated data are required, which usually demands expensive labelling resources and human effort to prepare the dataset. The biggest advantage of RNN and LSTMs is that they can model time series data (nearly all sensor data) and temporal relationships very well. Additionally, RNN and LSTMs can accept flexible input data size. The factors that prevent RNN and LSTMs from becoming the de facto method in DL-based HAR is that they are difficult to train in multiple aspects. They require a long training time and are very susceptible to diminishing/exploding gradients. It is also difficult to train them to efficiently model long time series.

GAN, as a generative model, can be used as a data augmentation method. Because it has a strong expressive capability to learn and imitate the latent data distribution of the targeted data, it outperforms traditional data augmentation methods [36]. Owing to its inherent data augmentation ability, GAN has the advantage of alleviating data demands at the beginning. However, GAN is often considered as hard to train because it alternatively trains a generator and a discriminator. Many variants of GAN and special training techniques have been proposed to tackle the converging issue [257–259].

Reinforcement learning is a relatively new area that is being explored for select areas in HAR, such as modelling muscle and arm movements [230,231]. Reinforcement learning is a type of unsupervised learning because it does not require explicit labels. Additionally, due to its online nature, reinforcement learning agents can be trained online while deployed in a real system. However, reinforcement learning agents are often difficult and time-consuming to train. Additionally, in the realm of DL-based HAR, the reward of the agent has to be given by a human, as in the case of [230,232]. In other words, even though people do not have to give explicit labels, humans are still required to provide something akin to a label (the reward) to train the agent.

When starting to choose a DL approach, we have a list of factors to consider, including the complexity of the target problem, dataset size, the availability and size of annotation, data quality, available computing resource, as well as the requirement of training time. Firstly, we have to evaluate and examine the problem complexity to decide upon promising

venues of machine learning methods. For example, if the problem is simple enough to resolve with the provided sensor modality, it's very likely that manual feature engineering and traditional machine learning method can provide satisfying results thus no DL method is needed. Secondly, before we choose the routine of DL, we would like to make sure the dataset size is sufficient to support a DL method. The lack of a sufficiently large corpus of labelled high-quality data is a major reason why DL methods cannot produce an expected result. Normally, when training a DL model with a limited dataset size, the model will be prone to overfitting, and the generalizability will be sacrificed, thus using a very deep network may not be a good choice. One option is to go for a shallow neural network or a traditional ML approach. Another option is to utilize specific algorithms to make the most out of the data. To be specific, data augmentation methods such as GAN can be readily implemented. Thirdly, another determining factor is the availability and size of annotation. When there is a large corpus of unlabeled sensor data at hand, a semi-supervised learning scheme is a promising direction one could consider, which will be discussed later in this work. Besides the availability of sensor data, the data quality also influences the network design. If the sensor is vulnerable to environmental noise, inducing a small SNR, some type of denoising structure (e.g., denoising auto-encoder) and increasing depth of the model can be considered to increase the noise-resiliency of the DL model. At last, a full evaluation of available computing resources and expected model training time cannot be more important for developers and researchers to choose a suitable DL approach.

6. Challenges and Opportunities

Though HAR has seen rapid growth, there are still a number of challenges that, if addressed, could further improve the status quo, leading to increased adoption of novel HAR techniques in existing and future wearables. In this section, we discuss these challenges and opportunities in HAR. Note that the issues discussed here are applicable to general HAR, not only DL-based HAR. We look to discuss and analyze the following four questions under our research question Q3 (challenges and opportunities), which overlap with the four major constituents of machine learning.

- Q3.1 : What are the challenges in data acquisition? How do we resolve them?
- Q3.2 : What are the challenges in label acquisition? What are the current methods?
- Q3.3 : What are the challenges in modeling? What are potential solutions?
- Q3.4 : What are the challenges in model deployment? What are potential opportunities?

6.1. Challenges in Data Acquisition

Data is the cornerstone of artificial intelligence. Models only perform as well as the quality of the training data. To build generalizable models, careful attention should be paid to data collection, ensuring the participants are representative of the population of interest. Moreover, determining a sufficient training dataset size is important in HAR. Currently, there is no well-defined method for determining the sample size of training data. However, showing the convergence of the error rate as a function of training data size is one approach shown by Yang et al. [260]. Acquiring a massive amount of high-quality data at a low cost is critical in every domain. In HAR, collecting raw data is labor-intensive considering a large number of different wearables. Therefore, proposing and developing innovative approaches to augmenting data with high quality is imperative for the growth of HAR research.

6.1.1. The Need for More Data

Data collection requires a considerable amount of effort in HAR. Particularly when researchers propose their original hardware, it is inevitable to collect data on users. Data augmentation is commonly used to generate synthetic training data when there is a data shortage. Synthetic noise is applied to real data to obtain new training samples. In general, using the dataset augmented with synthetic training samples yields higher classification accuracy when compared to using the original dataset [36,56,261,262]. Giorgi et al.

augmented their dataset by varying each signal sample with translation drawn from a small uniform distribution and showed improvements in accuracy using this augmented dataset [56]. Ismail Fawaz et al. [262] utilized Dynamic Time Warping to augment data and tested on UCR archive [105]. Deep learning methods are also used to augment the datasets to improve performance [238,263,264]. Alzantot et al. [264] and Wang et al. [238] employed GAN to synthesize sensor data using existing sensor data. Ramponi et al. [263] designed a conditional GAN-based framework to generate new irregularly-sampled time series to augment unbalanced data sets. Several works extracted 3D motion information from videos and transferred the knowledge to synthesize virtual on-body IMU sensor data [265,266]. In this way, they realized cross-modal IMU sensor data generation using traditional computer vision and graphics methods. **Opportunity:** We have listed some of the most recent works focusing on cross-modal sensor data synthesis. However, few researchers (if any) used a deep generative model to build a video-sensor multi-modal system. If we take a broader view, many works are using cross-modal deep generative models (such as GAN) in data synthesis, such as from video to audio [267], from text to image and vice versa [268,269]. Therefore, taking advantage of the cutting-edge deep generative models may contribute to addressing the wearable sensor data scarcity issue [270]. Another avenue of research is to utilize transfer learning, borrowing well-trained models from domains with high performing classifiers (i.e., images), and adapting them using a few samples of sensor data.

6.1.2. Data Quality and Missing Data

The quality of models is highly dependent on the quality of the training data. Many real-world collection scenarios introduce different sources of noise that degrade data quality, such as electromagnetic interference or uncertainty in task scheduling for devices that perform sampling [271]. In addition to improving hardware systems, multiple algorithms have been proposed to clean or impute poor-quality data. Data imputation is one of the most common methods to replace poor quality data or fill in missing data when sampling rates fluctuate greatly. For example, Cao et al. introduced a bi-directional recurrent neural network to impute time series data on the UCI localization dataset [272]. Luo et al. utilized a GAN to infer missing time series data [273]. Saeed et al. proposed an adversarial autoencoder (AAE) framework to perform data imputation [132]. **Opportunity:** To address this challenge, more research into automated methods for evaluating and quantifying the quality is needed to identify better, remove, and/or correct for poor quality data. Additionally, it has been experimentally shown that deep neural networks have the ability to learn well even if trained with noisy data, given that the networks are large enough and the dataset is large enough [274]. This motivates the need for HAR researchers to focus on other areas of importance, such as how to deploy larger models in real systems efficiently (Section 6.4) and generate more data (Section 6.1.1), which could potentially aid in solving this problem.

6.1.3. Privacy Protection

The privacy issue has become a concern among users [13]. In general, the more inference potential a sensor has, the less willing a person is to agree to its data collection. Multiple works have proposed privacy preservation methods while classifying human activities, including the replacement auto-encoder, the guardian, estimator, and neutralizer (GEN) architecture [128], and the anonymizing autoencoder [129]. For example, replacement auto-encoders learn to replace features of time-series data that correspond to sensitive inferences with values that correspond to non-sensitive inferences. Ultimately, these works obfuscate features that can identify the individual while preserving features common to each activity or movement. Federated learning is a trending approach to resolve privacy issues in learning problems [275–278]. It can enable the collaborative learning of a global model without the need to expose users' raw data. Xiao et al. [279] realized a federated averaging method combined with a perceptive extraction network to improve the performance of the federated learning system. Tu et al. [280] designed a dynamic layer sharing

scheme, which assisted the merging of local models to speed up the model convergence and achieved dynamic aggregation of models. Bettini et al. [281] presented a personalized semi-supervised federated learning method that built a global activity model and leveraged transfer learning for user personalization. Besides, Gudur and Perepu [282] implemented on-device federated learning using model distillation update and so-called weighted α-updates strategies to resolve model heterogeneities on a resource-limited embedded system (Raspberry Pi), which proved its effectiveness and efficiency. **Opportunity:** Blockchain is a new hot topic around the world. Blockchain, as a peer-to-peer network without the need for centralized authority, has been explored to facilitate the privacy-preserving data collection and sharing [283–286]. The combination of federated learning and blockchain is also a potential solution towards privacy protection [287] and is currently still in its very early stage. More collaboration between ubiquitous computing community and networking community should be encouraged to prosper in-depth research in novel directions.

6.2. Challenges in Label Acquisition

Labelled data is crucial for deep supervised learning. Image and audio data is generally easy to label by visual or aural confirmation. However, labelling human activities by looking at time series from HAR sensors is difficult or even impossible. Therefore, label acquisition for HAR sensors generally requires additional sensing sources to provide video or audio data to determine the ground truth, making label acquisition for HAR more labor-intensive. Moreover, accurate time synchronization between wearables and video/audio devices is challenging because different devices are equipped with independent (and often drifting) clocks. Several attempts have been made to address this issue, such as SyncWISE [288,289]. Two areas that require more research by the DL-HAR community are shortage in labelled data and difficulty in obtaining data from real-world scenarios.

6.2.1. Shortage of Labeled Data

As annotating large quantities of data is expensive, there have been great efforts to develop various methods to reduce the need for annotation, including data augmentation, semi-supervised learning, weakly supervised learning, and active learning to overcome this challenge. Semi-supervised learning utilizes both labelled data and unlabeled data to learn more generalizable feature representations. Zeng et al. presented two semi-supervised CNN methods that utilize unlabeled data during training: The convolutional encoder-decoder and the convolutional ladder network [290] and showed an 18% higher F1-score using the convolutional ladder network on the ActiTracker dataset. Dmitrijs demonstrated on the SHL dataset, with a CNN and AAE architecture, that semi-supervised learning on unlabeled data could achieve high accuracy [134]. Chen et al. proposed an encoder-decoder-based method that reduces distribution discrepancies between labelled and unlabeled data that arise due to differences in biology and behavior from different people while preserving the inherent similarities of different people performing the same task [291].

Active learning is a special type of semi-supervised learning that selectively chooses unlabeled data based on an objective function that selects data with low prediction confidence for a human annotator to label. Recently, researchers have tried to combine DL approaches with active learning to benefit from establishing labels on the fly while leveraging the extraordinary classification capability of DL. Gudur et al. utilized active learning by combining a CNN with Bayesian techniques to represent model uncertainties (B-CNN) [292]. Bettini et al. combined active learning and federated learning to proactively annotate the unlabeled sensor data and build personalized models in order to cope with data scarcity problem [281]. **Opportunity:** Though active learning has demonstrated that fewer labels are needed to build an effective deep neural network model, a real-world study with time-cost analysis would better demonstrate the benefits of active learning. Moreover, given the many existing labelled datasets, another area of opportunity is developing methods that leverage characteristics of labelled datasets to generate labels for unlabeled datasets such as transfer learning or pseudo-label method [293].

6.2.2. Issues of In-the-Field Dataset

Traditionally, HAR research has been conducted primarily in lab. Recently, HAR research has been moving towards in-field experiments. Unlike in-lab settings, where the ground truth can be captured by surveillance cameras, in-field experiments may have subjects moving around in daily life, where static camera deployment is not sufficient any more. Alharbi et al. used wearable cameras placed at the wrist, chest, and shoulder to record subject's activities as they moved around outside of a lab setting [294] and studied the feasibility of wearable cameras. **Opportunity:** More research in leveraging human-in-the-loop to provide in-field labelling is required to generate more robust datasets for in situ activities. Besides, one possible solution is to utilize existing in-the-field human activity video datasets and cross-modal deep generative models. If high-fidelity synthetic wearable sensor data can be generated from the available real-world video datasets (such as Stanford-ECM dataset [295]) or online video corpus, it may help alleviate the in-the-field data scarcity issue. Additionally, there are opportunities for semi-supervised learning methods that leverage the sparse labels provided by humans-in-the-loop to generate high-quality labels for the rest of the dataset.

6.3. Challenges in Modeling

In this section, we discuss the challenges and opportunities in the modelling process in several aspects, including data segmentation, semantically complex activity recognition, model generalizability, as well as model robustness.

6.3.1. Data Segmentation

As discussed in [296], many methods segment time series using traditional static sliding window methods. A static time window may either be too large, capturing more than necessary to detect certain activities, or too small and not capturing enough series to detect long movements. Recently, researchers have been looking to segment time series data more optimally. Zhang et al. used reinforcement learning to find more optimal activity segments to boost HAR performance [249]. Qian et al. [297] proposed weakly-supervised sensor-based activity segmentation and recognition method. **Opportunity:** More experimentation and research into dynamic activity segments or methods that leverage both short term and long term features (i.e., wavelets) are needed to create robust models at all timescales. While neural networks such as RNNs and LSTMs can model time series data with flexible time scales and automatically learn relevant features, their inherent issues such as exploding/vanishing gradients and training difficulty, make widespread adoption difficult. As such, more research into other methods that account for these issues is necessary.

6.3.2. Semantically Complex Activity Recognition

Current HAR methods achieve high performance for simple activities such as running. However, complex activities such as eating, which can involve a variety of movements, remain difficult. To tackle this challenge, Kyritsis et al. break down complex gestures into a series of simpler (atomic) gestures that, when combined, form the complex gesture [298]. Liu et al. propose a hierarchical architecture that constructs high-level human activities from low-level activities [299]. Peng et al. proposes AROMA, a complex human activity recognition method that leverages deep multi-task learning to learn simple activities that make up more complex movements [300]. **Opportunity:** Though hierarchical methods have been introduced for various complex tasks, there are still opportunities for improvements. Additionally, novel black-box approaches to complex task recognition, where individual steps in complex actions are automatically learned and accounted for rather than specifically identified or labelled by designers, have yet to be fully explored. Such a paradigm is perfectly suitable for deep learning because neural networks function on a similar principle. Besides, graph neural network can also be explored to model the hierarchical structure of simple-to-complex human activities [301].

6.3.3. Model Generalizability

A model has high generalizability when it performs well on data that it has never seen before. Overfitting occurs when it performs well on training data but poorly on new data. Recently, many efforts have been put into improving the generalizability of models in HAR [86,302,303]. Most research on generalizability in HAR has been focused on creating models that can generalize to a larger population, which often requires a large amount of data and high model complexity. In scenarios where high model complexity and data are not bottlenecks, DL-based HAR generally outperforms and generalize better than other types of methods. In scenarios where data or model complexity is limited, DL-based methods must utilize available data more efficiently or adapt to the specific scenario online. For instance, Siirtola and Röning propose an online incremental learning approach that continuously adapts the model with the user's individual data as it comes in [304]. Qian et al. [305] introduce Generalizable Independent Latent Excitation (GILE), which greatly enhances the cross-person generalization capability of the model. **Opportunity:** An avenue of generalizability that has yet to be fully explored are new training methods that can adapt and learn predictors across multiple environments, such as invariant risk minimization [306] or federated learning methods [307]. Incorporating these areas into DL-based HAR could not only improve the generalizability of HAR models but accomplish this in a model-agnostic way.

6.3.4. Model Robustness

A key issue that the community is paying increasing attention to is model robustness and reliability [308,309]. One common way to improve robustness is to leverage the benefits of multiple types of sensors together to create multi-sensory systems [249,310–314]. Huynh-The et al. [311] has proposed an architecture called DeepFusionHAR to incorporate the handcrafted features and deep learning extracted features from multiple sensors to detect daily life and sports activities. Hanif et al. [312] proposed a multi-sensory approach for basic and complex human activity recognition that uses built-in sensors from smartphones and smartwatches to classify 20 complex actions and five basic actions. Pires et al. [313] demonstrated a mobile application on a multi-sensor mobile platform for daily living activity classification using a combination of accelerometer, gyroscope, magnetometer, microphone, and GPS. Multi-sensory networks in some cases are integrated with attention modules to learn the most representative and discriminative sensor modality to distinguish human activities [249]. **Opportunity:** While there are works that utilize multiple sensors to improve robustness, they require users to wear or have access to all of the sensors they utilize. An exciting new direction is to create generalized frameworks that can adaptively utilize data from whatever sensors happen to be available, such as a smart home intelligence system [315]. For this direction, deep learning methods seem more suitable than classical machine learning methods because neural networks can be more easily tuned and adapted to different domains (i.e., different sensors) than rigid classical models, just by tuning weights or by mixing and matching different layers or embeddings. Creating such systems would not only greatly improve the practicality of HAR-based systems but would also contribute significantly to general artificial intelligence.

6.4. Challenges in Model Deployment

There are several works focusing on deploying deep-learning-based HAR on mobile platforms. Lane et al. [316] proposes a SOC-based architecture for optimizing deep neural network inference, while Lane et al. [317] and Cao et al. [318] utilize the smartphone's digital signal processor (DSP) and mobile GPU to improve inference time and reduce power consumption. Yao et al. [319] propose a lightweight CNN and RNN-based system that accounts for noisy sensor readings from smartphones and automatically learns local and global features between sensor windows to improve performance.

The second class of works focus on reducing the complexity of neural networks so that they can run on resource-limited mobile platforms. Bhattacharya and Lane [320] reduces the

amount of computation required at each layer by encoding layers into a lower-dimensional space. Edel and Köppe [321] reduces computation by utilizing binary weights rather than fixed-point or floating weights.

Emerging trends in deploying neural networks include offloading computation onto application-specific integrated circuits (ASIC) or lower power consumption microcontrollers. Bhat et al. [322], Wang et al. [323] developed custom integrated circuits and hardware accelerators that perform the entire HAR pipeline with significantly lower power consumption than mobile or GPU-based platforms. The downside to ASICs is that they cannot be reconfigured for other types of tasks. Islam and Nirjon [324] present an architecture for embedded systems that dynamically schedules DNN inference tasks to improve inference time and accuracy.

Opportunity: Though there are works that explore the deployment of DNNs practical systems, more research is needed for society to fully benefit from the advances in DNNs for HAR. Many of the works discussed leverage a single platform (i.e., either a smartphone or ASIC), but there are still many opportunities for improving the practical use of HAR by exploring intelligent ways to partition computation across the cloud, mobile platforms, and other edge devices. DNN-based HAR systems can largely benefit by incorporating methodologies proposed by works such as [325–332], that carefully partition computation and data across multiple devices and the cloud.

Lane et al. performed a small-scale exploration into the performance of DNNs for HAR applications on mobile platforms in various configurations, including utilizing the phone's CPU and DSP and offloading computation onto remote devices [333]. This work demonstrates that mobile devices running DNN inference can scale gracefully across different compute resources available to the mobile platform and also supports the need for more research into optimal strategies for partitioning DNN inference across mobile and edge systems to improve latency, reduce power consumption, and increase the complexity of the DNNs serviceable to wearable platforms.

7. Conclusions

Human activity recognition in wearables has provided us with many conveniences and avenues to monitor and improve our life quality. AI and ML have played a vital role in enabling HAR in wearables. In recent years, DL has pushed the boundary of wearables-based HAR, bringing activity recognition performance to an all-time high. In this paper, we provided our answers to the three research questions we proposed in Section 2. We firstly gave an overall picture of the real-life applications, mainstream sensors, and popular public datasets of HAR. Then we gave a review of the advances of the deep learning approaches used in the field of wearable HAR and provided guidelines and insights about how to choose an appropriate DL approach after comparing the advantages and disadvantages of them. At last, we discussed the current road blockers in three aspects—data-wise, label-wise, and model-wise—for each of which we provide potential opportunities. We further identify the open challenges and finally provide suggestions for future avenues of research in this field. By categorizing and summarizing existing works that apply DL approaches to wearable sensor-based HAR, we aim to provide new engineers and researchers entering this field an overall picture of the existing research work and remaining challenges. We would also like to benefit experienced researchers by analyzing and discussing the developing trends, major barriers, cutting-edge frontiers, and potential future directions.

References

1. Vogels, E.A. About One-in-five Americans Use a Smart Watch or Fitness Tracker. Available online: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ (accessed on 10 February 2022).
2. Research, M. Wearable Devices Market by Product Type (Smartwatch, Earwear, Eyewear, and others), End-Use Industry (Consumer Electronics, Healthcare, Enterprise and Industrial, Media and Entertainment), Connectivity Medium, and Region—Global Forecast to 2025. Available online: https://www.meticulousresearch.com/product/wearable-devices-market-5050 (accessed on 10 February 2022).
3. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.* **1989**, *2*, 303–314. [CrossRef]
4. Schäfer, A.M.; Zimmermann, H.G. Recurrent Neural Networks Are Universal Approximators. In *Artificial Neural Networks—ICANN 2006*; Kollias, S.D., Stafylopatis, A., Duch, W., Oja, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 632–640.
5. Zhou, D.X. Universality of deep convolutional neural networks. *Appl. Comput. Harmon. Anal.* **2020**, *48*, 787–794. [CrossRef]
6. Wearable Technology Database. Available online: https://data.world/crowdflower/wearable-technology-database (accessed on 10 February 2022).
7. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning. 2016. Available online: http://www.deeplearningbook.org (accessed on 10 February 2022).
8. Sutton, R.S.; Barto, A.G. Reinforcement Learning: An Introduction. 2018. Available online: http://www.incompleteideas.net/book/the-book-2nd.html (accessed on 10 February 2022).
9. Transparent Reporting of Systematic Reviews and Meta-Analyses. Available online: http://www.prisma-statement.org/ (accessed on 10 February 2022).
10. Kiran, S.; Khan, M.A.; Javed, M.Y.; Alhaisoni, M.; Tariq, U.; Nam, Y.; Damasevicius, R.; Sharif, M. Multi-Layered Deep Learning Features Fusion for Human Action Recognition. *Comput. Mater. Contin.* **2021**, *69*, 4061–4075. [CrossRef]
11. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [CrossRef]
12. Nweke, H.F.; Teh, Y.W.; Al-Garadi, M.A.; Alo, U.R. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst. Appl.* **2018**, *105*, 233–261. [CrossRef]
13. Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; Liu, Y. Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities. *ACM Comput. Surv.* **2021**, *54*, 1–40. [CrossRef]
14. Ramanujam, E.; Perumal, T.; Padmavathi, S. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sens. J.* **2021**, *21*, 13029–13040. [CrossRef]
15. Morales, J.; Akopian, D. Physical activity recognition by smartphones, a survey. *Biocybern. Biomed. Eng.* **2017**, *37*, 388–400. [CrossRef]
16. Booth, F.W.; Roberts, C.K.; Laye, M.J. Lack of exercise is a major cause of chronic diseases. *Compr. Physiol.* **2011**, *2*, 1143–1211.
17. Bauman, A.E.; Reis, R.S.; Sallis, J.F.; Wells, J.C.; Loos, R.J.; Martin, B.W. Correlates of physical activity: Why are some people physically active and others not? *Lancet* **2012**, *380*, 258–271. [CrossRef]
18. Diaz, K.M.; Krupka, D.J.; Chang, M.J.; Peacock, J.; Ma, Y.; Goldsmith, J.; Schwartz, J.E.; Davidson, K.W. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *Int. J. Cardiol.* **2015**, *185*, 138–140. [CrossRef] [PubMed]
19. Zhu, J.; Pande, A.; Mohapatra, P.; Han, J.J. Using Deep Learning for Energy Expenditure Estimation with wearable sensors. In Proceedings of the 2015 17th International Conference on E-health Networking, Application Services (HealthCom), Boston, MA, USA, 14–17 October 2015; pp. 501–506. [CrossRef]
20. Brown, V.; Moodie, M.; Herrera, A.M.; Veerman, J.; Carter, R. Active transport and obesity prevention–a transportation sector obesity impact scoping review and assessment for Melbourne, Australia. *Prev. Med.* **2017**, *96*, 49–66. [CrossRef] [PubMed]
21. Bisson, A.; Lachman, M.E. Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity. *Front. Public Health* **2017**, *4*, 289. [CrossRef]
22. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional Neural Networks for human activity recognition using mobile sensors. In Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, Austin, TX, USA, 6–7 November 2014; pp. 197–205. [CrossRef]
23. Chen, Y.; Xue, Y. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 1488–1492. [CrossRef]

24. Jiang, W.; Yin, Z. Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks. In Proceedings of the 23rd ACM International Conference on Multimedia (MM), Brisbane, Australia, 26–30 October 2015; ACM: New York, NY, USA, 2015; pp. 1307–1310. [CrossRef]

25. Ronao, C.A.; Cho, S.B. Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks. *Expert Syst. Appl.* **2016**, *59*, 235–244. [CrossRef]

26. Lee, S.M.; Yoon, S.M.; Cho, H. Human activity recognition from accelerometer data using Convolutional Neural Network. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea, 13–16 February 2017; pp. 131–134. [CrossRef]

27. Wang, L.; Gjoreski, H.; Ciliberto, M.; Mekki, S.; Valentin, S.; Roggen, D. Benchmarking the SHL Recognition Challenge with Classical and Deep-Learning Pipelines. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp), Singapore, 8–12 October 2018; ACM: New York, NY, USA, 2018; pp. 1626–1635. [CrossRef]

28. Li, S.; Li, C.; Li, W.; Hou, Y.; Cook, C. Smartphone-sensors Based Activity Recognition Using IndRNN. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, (UbiComp), Singapore, 8–12 October 2018; ACM: New York, NY, USA, 2018; pp. 1541–1547. [CrossRef]

29. Jeyakumar, J.V.; Lee, E.S.; Xia, Z.; Sandha, S.S.; Tausik, N.; Srivastava, M. Deep Convolutional Bidirectional LSTM Based Transportation Mode Recognition. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp), Singapore, 8–12 October 2018; ACM: New York, NY, USA, 2018; pp. 1606–1615. [CrossRef]

30. Wang, K.; He, J.; Zhang, L. Attention-based Convolutional Neural Network for Weakly Labeled Human Activities Recognition with Wearable Sensors. *IEEE Sens. J.* **2019**, *19*, 7598–7604. [CrossRef]

31. Hammerla, N.Y.; Fisher, J.M.; Andras, P.; Rochester, L.; Walker, R.; Plotz, T. PD Disease State Assessment in Naturalistic Environments Using Deep Learning. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI), Hyatt Regency, Austin, TX, USA, 25–30 January 2015; pp. 1742–1748.

32. Eskofier, B.M.; Lee, S.I.; Daneault, J.; Golabchi, F.N.; Ferreira-Carvalho, G.; Vergara-Diaz, G.; Sapienza, S.; Costante, G.; Klucken, J.; Kautz, T.; et al. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Lake Buena Vista (Orlando), FL, USA, 16–20 August 2016; pp. 655–658. [CrossRef]

33. Zhang, A.; Cebulla, A.; Panev, S.; Hodgins, J.; De la Torre, F. Weakly-supervised learning for Parkinson's Disease tremor detection. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 143–147. [CrossRef]

34. Mohammadian Rad, N.; Van Laarhoven, T.; Furlanello, C.; Marchiori, E. Novelty Detection Using Deep Normative Modeling for IMU-Based Abnormal Movement Monitoring in Parkinson's Disease and Autism Spectrum Disorders. *Sensors* **2018**, *18*, 3533. [CrossRef]

35. Kim, H.B.; Lee, W.W.; Kim, A.; Lee, H.J.; Park, H.Y.; Jeon, H.S.; Kim, S.K.; Jeon, B.; Park, K.S. Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network. *Comput. Biol. Med.* **2018**, *95*, 140–146. [CrossRef]

36. Um, T.T.; Pfister, F.M.J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI), Glasgow, UK, 13–17 November 2017; ACM: New York, NY, USA, 2017; pp. 216–220. [CrossRef]

37. Xu, X.; Nemati, E.; Vatanparvar, K.; Nathan, V.; Ahmed, T.; Rahman, M.M.; McCaffrey, D.; Kuang, J.; Gao, J.A. Listen2Cough: Leveraging End-to-End Deep Learning Cough Detection Model to Enhance Lung Health Assessment Using Passively Sensed Audio. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–22. [CrossRef]

38. Zhang, S.; Nemati, E.; Ahmed, T.; Rahman, M.M.; Kuang, J.; Gao, A. A Novel Multi-Centroid Template Matching Algorithm and Its Application to Cough Detection. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Guadalajara, Mexico, 31 October–4 November 2021; pp. 7598–7604. [CrossRef]

39. Nemati, E.; Zhang, S.; Ahmed, T.; Rahman, M.M.; Kuang, J.; Gao, A. CoughBuddy: Multi-Modal Cough Event Detection Using Earbuds Platform. In Proceedings of the 2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Athens, Greece, 27–30 July 2021; pp. 1–4. [CrossRef]

40. Zhang, S.; Nemati, E.; Dinh, M.; Folkman, N.; Ahmed, T.; Rahman, M.; Kuang, J.; Alshurafa, N.; Gao, A. CoughTrigger: Earbuds IMU Based Cough Detection Activator Using An Energy-efficient Sensitivity-prioritized Time Series Classifier. *arXiv* **2021**, arXiv:2111.04185.

41. Gao, Y.; Long, Y.; Guan, Y.; Basu, A.; Baggaley, J.; Ploetz, T. Towards Reliable, Automated General Movement Assessment for Perinatal Stroke Screening in Infants Using Wearable Accelerometers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 12:1–12:22. [CrossRef]

42. Ghandeharioun, A.; Fedor, S.; Sangermano, L.; Ionescu, D.; Alpert, J.; Dale, C.; Sontag, D.; Picard, R. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 325–332. [CrossRef]

43. Phinyomark, A.; Scheme, E. EMG Pattern Recognition in the Era of Big Data and Deep Learning. *Big Data Cogn. Comput.* **2018**, *2*, 21. [CrossRef]

44. Meattini, R.; Benatti, S.; Scarcia, U.; De Gregorio, D.; Benini, L.; Melchiorri, C. An sEMG-Based Human–Robot Interface for Robotic Hands Using Machine Learning and Synergies. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2018**, *8*, 1149–1158. [CrossRef]

45. Parajuli, N.; Sreenivasan, N.; Bifulco, P.; Cesarelli, M.; Savino, S.; Niola, V.; Esposito, D.; Hamilton, T.J.; Naik, G.R.; Gunawardana, U.; et al. Real-time EMG based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation. *Sensors* **2019**, *19*, 4596. [CrossRef] [PubMed]

46. Samuel, O.W.; Asogbon, M.G.; Geng, Y.; Al-Timemy, A.H.; Pirbhulal, S.; Ji, N.; Chen, S.; Fang, P.; Li, G. Intelligent EMG pattern recognition control method for upper-limb multifunctional prostheses: Advances, current challenges, and future prospects. *IEEE Access* **2019**, *7*, 10150–10165. [CrossRef]

47. Zhao, H.; Ma, Y.; Wang, S.; Watson, A.; Zhou, G. MobiGesture: Mobility-aware hand gesture recognition for healthcare. *Smart Health* **2018**, *9–10*, 129–143. [CrossRef]

48. Shin, S.; Sung, W. Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks. In Proceedings of the 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, Canada, 22–25 May 2016; pp. 2274–2277. [CrossRef]

49. Zhang, C.; Xue, Q.; Waghmare, A.; Meng, R.; Jain, S.; Han, Y.; Li, X.; Cunefare, K.; Ploetz, T.; Starner, T.; et al. FingerPing: Recognizing Fine-grained Hand Poses Using Active Acoustic On-body Sensing. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18), Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 437:1–437:10. [CrossRef]

50. Pacchierotti, C.; Salvietti, G.; Hussain, I.; Meli, L.; Prattichizzo, D. The hRing: A wearable haptic device to avoid occlusions in hand tracking. In Proceedings of the 2016 IEEE Haptics Symposium (HAPTICS), Philadelphia, PA, USA, 8–11 April 2016; pp. 134–139. [CrossRef]

51. Sundaram, S.; Kellnhofer, P.; Li, Y.; Zhu, J.Y.; Torralba, A.; Matusik, W. Learning the signatures of the human grasp using a scalable tactile glove. *Nature* **2019**, *569*, 698–702. [CrossRef]

52. Kim, J.; Kim, M.; Kim, K. Development of a wearable HCI controller through sEMG & IMU sensor fusion. In Proceedings of the 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Sofitel Xian on Renmin Square, Xi'an, China, 19–22 August 2016; pp. 83–87. [CrossRef]

53. Kalantarian, H.; Alshurafa, N.; Le, T.; Sarrafzadeh, M. Monitoring eating habits using a piezoelectric sensor-based necklace. *Comput. Biol. Med.* **2015**, *58*, 46–55. [CrossRef]

54. Zhang, S.; Zhao, Y.; Nguyen, D.T.; Xu, R.; Sen, S.; Hester, J.; Alshurafa, N. NeckSense: A Multi-Sensor Necklace for Detecting Eating Activities in Free-Living Conditions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 1–26. [CrossRef]

55. Chen, T.; Li, Y.; Tao, S.; Lim, H.; Sakashita, M.; Zhang, R.; Guimbretiere, F.; Zhang, C. NeckFace: Continuously Tracking Full Facial Expressions on Neck-Mounted Wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–31. [CrossRef]

56. Giorgi, G.; Martinelli, F.; Saracino, A.; Sheikhalishahi, M. Try Walking in My Shoes, if You Can: Accurate Gait Recognition Through Deep Learning. In *Computer Safety, Reliability, and Security*; Tonetta, S., Schoitsch, E., Bitsch, F., Eds.; Springer: Cham, Switzerland, 2017; pp. 384–395.

57. Shi, Y.; Manco, M.; Moyal, D.; Huppert, G.; Araki, H.; Banks, A.; Joshi, H.; McKenzie, R.; Seewald, A.; Griffin, G.; et al. Soft, stretchable, epidermal sensor with integrated electronics and photochemistry for measuring personal UV exposures. *PLoS ONE* **2018**, *13*, 1–15. [CrossRef]

58. Chung, S.; Lim, J.; Noh, K.J.; Gue Kim, G.; Jeong, H.T. Sensor Positioning and Data Acquisition for Activity Recognition using Deep Learning. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 17–19 October 2018; pp. 154–159. [CrossRef]

59. Laput, G.; Xiao, R.; Harrison, C. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST'16), Tokyo, Japan, 16–19 October 2016; ACM: New York, NY, USA, 2016; pp. 321–333. [CrossRef]

60. Laput, G.; Harrison, C. Sensing Fine-Grained Hand Activity with Smartwatches. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19), Glasgow, UK, 4-9 May, 2019; ACM: New York, NY, USA, 2019; pp. 338:1–338:13. [CrossRef]

61. Electromyography. Electromyography—Wikipedia, The Free Encyclopedia. 2010. Available online: https://en.wikipedia.org/wiki/Electromyography (accessed on 10 March 2020).

62. Zia ur Rehman, M.; Waris, A.; Gilani, S.O.; Jochumsen, M.; Niazi, I.K.; Jamil, M.; Farina, D.; Kamavuako, E.N. Multiday EMG-based classification of hand motions with deep learning techniques. *Sensors* **2018**, *18*, 2497. [CrossRef] [PubMed]

63. Triwiyanto, T.; Pawana, I.P.A.; Purnomo, M.H. An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1678–1688. [CrossRef] [PubMed]

64. Wu, H.; Huang, Q.; Wang, D.; Gao, L. A CNN-SVM combined model for pattern recognition of knee motion using mechanomyography signals. *J. Electromyogr. Kinesiol.* **2018**, *42*, 136–142. [CrossRef] [PubMed]

65. Meagher, C.; Franco, E.; Turk, R.; Wilson, S.; Steadman, N.; McNicholas, L.; Vaidyanathan, R.; Burridge, J.; Stokes, M. New advances in mechanomyography sensor technology and signal processing: Validity and intrarater reliability of recordings from muscle. *J. Rehabil. Assist. Technol. Eng.* **2020**, *7*, 2055668320916116. [CrossRef]

66. Khalifa, S.; Lan, G.; Hassan, M.; Seneviratne, A.; Das, S.K. Harke: Human activity recognition from kinetic energy harvesting data in wearable devices. *IEEE Trans. Mob. Comput.* **2017**, *17*, 1353–1368. [CrossRef]

67. Cha, Y.; Kim, H.; Kim, D. Flexible piezoelectric sensor-based gait recognition. *Sensors* **2018**, *18*, 468. [CrossRef]

68. Massé, F.; Gonzenbach, R.R.; Arami, A.; Paraschiv-Ionescu, A.; Luft, A.R.; Aminian, K. Improving activity recognition using a wearable barometric pressure sensor in mobility-impaired stroke patients. *J. Neuroeng. Rehabil.* **2015**, *12*, 1–15. [CrossRef]

69. Barna, A.; Masum, A.K.M.; Hossain, M.E.; Bahadur, E.H.; Alam, M.S. A study on human activity recognition using gyroscope, accelerometer, temperature and humidity data. In Proceedings of the 2019 International conference on electrical, computer and communication engineering (ECCE), Cox'sBazar, Bangladesh, 7–9 February 2019; pp. 1–6.

70. Salehzadeh, A.; Calitz, A.P.; Greyling, J. Human activity recognition using deep electroencephalography learning. *Biomed. Signal Process. Control* **2020**, *62*, 102094. [CrossRef]

71. Ramos-Garcia, R.I.; Tiffany, S.; Sazonov, E. Using respiratory signals for the recognition of human activities. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Lake Buena Vista (Orlando), FL, USA, 16–20 August 2016; pp. 173–176.

72. Filippoupolitis, A.; Takand, B.; Loukas, G. Activity recognition in a home setting using off the shelf smart watch technology. In Proceedings of the 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS), Granada, Spain, 14–16 December 2016; pp. 39–44.

73. Jin, Y.; Gao, Y.; Zhu, Y.; Wang, W.; Li, J.; Choi, S.; Li, Z.; Chauhan, J.; Dey, A.K.; Jin, Z. SonicASL: An Acoustic-based Sign Language Gesture Recognizer Using Earphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–30. [CrossRef]

74. Ntalampiras, S.; Potamitis, I. Transfer learning for improved audio-based human activity recognition. *Biosensors* **2018**, *8*, 60. [CrossRef]

75. Hamid, A.; Brahim, A.; Mohammed, O. A survey of activity recognition in egocentric lifelogging datasets. In Proceedings of the 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 19–20 April 2017; pp. 1–8.

76. Ryoo, M.S.; Matthies, L. First-Person Activity Recognition: What Are They Doing to Me? In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 2730–2737.

77. Alharbi, R.; Stump, T.; Vafaie, N.; Pfammatter, A.; Spring, B.; Alshurafa, N. I can't be myself: Effects of wearable cameras on the capture of authentic behavior in the wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–40. [CrossRef] [PubMed]

78. Alharbi, R.; Tolba, M.; Petito, L.C.; Hester, J.; Alshurafa, N. To Mask or Not to Mask? Balancing Privacy with Visual Confirmation Utility in Activity-Oriented Wearable Cameras. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–29. [CrossRef] [PubMed]

79. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef] [PubMed]

80. Xu, S.; Wang, J.; Shou, W.; Ngo, T.; Sadick, A.M.; Wang, X. Computer vision techniques in construction: A critical review. *Arch. Comput. Methods Eng.* **2021**, *28*, 3383–3397. [CrossRef]

81. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explor. Newsl.* **2011**, *12*, 74–82. [CrossRef]

82. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. In Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 24–26 April 2013.

83. van Kasteren, T.; Noulas, A.; Englebienne, G.; Kröse, B. Accurate Activity Recognition in a Home Setting. In Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp'08), Seoul, Korea, 21–24 September 2008; pp. 1–9. [CrossRef]

84. Shoaib, M.; Bosch, S.; Incel, O.; Scholten, H.; Havinga, P. Fusion of smartphone motion sensors for physical activity recognition. *Sensors* **2014**, *14*, 10146–10176. [CrossRef]

85. Chen, C.; Jafari, R.; Kehtarnavaz, N. UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 168–172. [CrossRef]

86. Stisen, A.; Blunck, H.; Bhattacharya, S.; Prentow, T.S.; Kjærgaard, M.B.; Dey, A.; Sonne, T.; Jensen, M.M. Smart Devices Are Different: Assessing and MitigatingMobile Sensing Heterogeneities for Activity Recognition. In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys'15), Seoul, Korea, 1–4 November 2015; pp. 127–140. [CrossRef]

87. Altun, K.; Barshan, B.; Tunçel, O. Comparative Study on Classifying Human Activities with Miniature Inertial and Magnetic Sensors. *Pattern Recogn.* **2010**, *43*, 3605–3620. [CrossRef]

88. Banos, O.; Garcia, R.; Holgado-Terriza, J.A.; Damas, M.; Pomares, H.; Rojas, I.; Saez, A.; Villalonga, C. mHealthDroid: A novel framework for agile development of mobile health applications. In *International Workshop on Ambient Assisted Living*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 91–98.

89. Banos, O.; Villalonga, C.; Garcia, R.; Saez, A.; Damas, M.; Holgado-Terriza, J.A.; Lee, S.; Pomares, H.; Rojas, I. Design, implementation and validation of a novel open framework for agile development of mobile health applications. *Biomed. Eng. Online* **2015**, *14*, S6. [CrossRef]

90. Chavarriaga, R.; Sagha, H.; Calatroni, A.; Digumarti, S.T.; Tröster, G.; Millán, J.D.R.; Roggen, D. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognit. Lett.* **2013**, *34*, 2033–2042. [CrossRef]

91. Reiss, A.; Stricker, D. Introducing a New Benchmarked Dataset for Activity Monitoring. In Proceedings of the 2012 16th International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012; pp. 108–109. [CrossRef]

92. Shoaib, M.; Scholten, H.; Havinga, P.J.M. Towards Physical Activity Recognition Using Smartphone Sensors. In Proceedings of the 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, DC, USA, 18–21 December 2013; pp. 80–87. [CrossRef]

93. Micucci, D.; Mobilio, M.; Napoletano, P. UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones. *Appl. Sci.* **2017**, *7*, 1101. [CrossRef]

94. Zhang, M.; Sawchuk, A.A. USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp'12), Pittsburgh, PA, USA, 5–8 September 2012; pp. 1036–1043. [CrossRef]

95. Vaizman, Y.; Ellis, K.; Lanckriet, G. Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches. *IEEE Pervasive Comput.* **2017**, *16*, 62–74. [CrossRef]

96. Kawaguchi, N.; Ogawa, N.; Iwasaki, Y.; Kaji, K.; Terada, T.; Murao, K.; Inoue, S.; Kawahara, Y.; Sumi, Y.; Nishio, N. HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-world Activity Understandings. In Proceedings of the 2nd Augmented Human International Conference, (AH'11), Tokyo, Japan, 13 March 2011; pp. 27:1–27:5. [CrossRef]

97. Weiss, G.M.; Lockhart, J.W.; Pulickal, T.T.; McHugh, P.T.; Ronan, I.H.; Timko, J.L. Actitracker: A Smartphone-Based Activity Recognition System for Improving Health and Well-Being. In Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada, 17–19 October 2016; pp. 682–688. [CrossRef]

98. Bruno, B.; Mastrogiovanni, F.; Sgorbissa, A. A public domain dataset for ADL recognition using wrist-placed accelerometers. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 738–743. [CrossRef]

99. Zhang, Z.; Pi, Z.; Liu, B. TROIKA: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmographic Signals During Intensive Physical Exercise. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 522–531. [CrossRef] [PubMed]

100. Bulling, A.; Blanke, U.; Schiele, B. A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors. *ACM Comput. Surv.* **2014**, *46*, 33:1–33:33. [CrossRef]

101. Kyritsis, K.; Tatli, C.L.; Diou, C.; Delopoulos, A. Automated analysis of in meal eating behavior using a commercial wristband IMU sensor. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 2843–2846. [CrossRef]

102. Chauhan, J.; Hu, Y.; Seneviratne, S.; Misra, A.; Seneviratne, A.; Lee, Y. BreathPrint: Breathing Acoustics-based User Authentication. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'17), Niagara Falls, NY, USA, 19–23 June 2017; pp. 278–291. [CrossRef]

103. Zappi, P.; Lombriser, C.; Stiefmeier, T.; Farella, E.; Roggen, D.; Benini, L.; Tröster, G. Activity Recognition from On-body Sensors: Accuracy-power Trade-off by Dynamic Sensor Selection. In Proceedings of the 5th European Conference on Wireless Sensor Networks (EWSN'08), Bologna, Italy, 30 January 2008–1 February 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 17–33.

104. Bachlin, M.; Plotnik, M.; Roggen, D.; Maidan, I.; Hausdorff, J.M.; Giladi, N.; Troster, G. Wearable Assistant for Parkinson's Disease Patients with the Freezing of Gait Symptom. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 436–446. [CrossRef] [PubMed]

105. Chen, Y.; Keogh, E.; Hu, B.; Begum, N.; Bagnall, A.; Mueen, A.; Batista, G. The UCR Time Series Classification Archive. 2015. Available online: www.cs.ucr.edu/~eamonn/time_series_data/ (accessed on 10 February 2022).

106. Bagnall, A.; Dau, H.A.; Lines, J.; Flynn, M.; Large, J.; Bostrom, A.; Southam, P.; Keogh, E. The UEA multivariate time series classification archive. *arXiv* **2018**, arXiv:1811.00075.

107. Liu, J.; Wang, Z.; Zhong, L.; Wickramasuriya, J.; Vasudevan, V. uWave: Accelerometer-based personalized gesture recognition and its applications. In Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications, Galveston, TX, USA, 9–13 March 2009; pp. 1–9. [CrossRef]

108. Gjoreski, H.; Ciliberto, M.; Wang, L.; Ordonez Morales, F.J.; Mekki, S.; Valentin, S.; Roggen, D. The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics with Mobile Devices. *IEEE Access* **2018**, *6*, 42592–42604. [CrossRef]

109. Dau, H.A.; Bagnall, A.; Kamgar, K.; Yeh, C.C.M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C.A.; Keogh, E. The UCR Time Series Archive. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 1293–1305. [CrossRef]

110. Siddiqui, N.; Chan, R.H.M. A wearable hand gesture recognition device based on acoustic measurements at wrist. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 4443–4446. [CrossRef]

111. Wang, H.; Li, L.; Chen, H.; Li, Y.; Qiu, S.; Gravina, R. Motion recognition for smart sports based on wearable inertial sensors. In *EAI International Conference on Body Area Networks*; Springer: Cham, Switzerland, 2019; pp. 114–124.

112. Hassan, M.M.; Uddin, M.Z.; Mohamed, A.; Almogren, A. A robust human activity recognition system using smartphone sensors and deep learning. *Future Gener. Comput. Syst.* **2018**, *81*, 307–313. [CrossRef]

113. Yang, J.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.

114. Moya Rueda, F.; Grzeszick, R.; Fink, G.A.; Feldhorst, S.; Ten Hompel, M. Convolutional neural networks for human activity recognition using body-worn sensors. *Informatics* **2018**, *5*, 26. [CrossRef]

115. Li, F.; Shirahama, K.; Nisar, M.A.; Köping, L.; Grzegorzek, M. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors* **2018**, *18*, 679. [CrossRef]

116. Kim, E. Interpretable and accurate convolutional neural networks for human activity recognition. *IEEE Trans. Ind. Informatics* **2020**, *16*, 7190–7198. [CrossRef]

117. Tang, Y.; Teng, Q.; Zhang, L.; Min, F.; He, J. Layer-Wise Training Convolutional Neural Networks With Smaller Filters for Human Activity Recognition Using Wearable Sensors. *IEEE Sens. J.* **2021**, *21*, 581–592. [CrossRef]

118. Sun, J.; Fu, Y.; Li, S.; He, J.; Xu, C.; Tan, L. Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. *J. Sens.* **2018**, *2018*, 8580959. [CrossRef]

119. Ballard, D.H. Modular Learning in Neural Networks. In Proceedings of the Sixth National Conference on Artificial Intelligence, AAAI'87, Washington, DC, USA, 13–17 July 1987; Volume 1, pp. 279–284.

120. Varamin, A.A.; Abbasnejad, E.; Shi, Q.; Ranasinghe, D.C.; Rezatofighi, H. Deep auto-set: A deep auto-encoder-set network for activity recognition using wearables. In Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, New York, NY, USA, 2–7 November 2018; pp. 246–253.

121. Malekzadeh, M.; Clegg, R.G.; Haddadi, H. Replacement autoencoder: A privacy-preserving algorithm for sensory data analysis. In Proceedings of the 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 17–20 April 2018; pp. 165–176.

122. Jia, G.; Lam, H.K.; Liao, J.; Wang, R. Classification of Electromyographic Hand Gesture Signals using Machine Learning Techniques. *Neurocomputing* **2020**, *401*, 236–248. [CrossRef]

123. Rubio-Solis, A.; Panoutsos, G.; Beltran-Perez, C.; Martinez-Hernandez, U. A multilayer interval type-2 fuzzy extreme learning machine for the recognition of walking activities and gait events using wearable sensors. *Neurocomputing* **2020**, *389*, 42–55. [CrossRef]

124. Gavrilin, Y.; Khan, A. Across-Sensor Feature Learning for Energy-Efficient Activity Recognition on Mobile Devices. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–7. [CrossRef]

125. Li, Y.; Shi, D.; Ding, B.; Liu, D. Unsupervised Feature Learning for Human Activity Recognition Using Smartphone Sensors. In *Mining Intelligence and Knowledge Exploration*; Prasath, R., O'Reilly, P., Kathirvalavakumar, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 99–107.

126. Almaslukh, B.; AlMuhtadi, J.; Artoli, A. An effective deep autoencoder approach for online smartphone-based human activity recognition. *Int. J. Comput. Sci. Netw. Secur.* **2017**, *17*, 160.

127. Mohammed, S.; Tashev, I. Unsupervised deep representation learning to remove motion artifacts in free-mode body sensor networks. In Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Eindhoven, The Netherlands, 9–12 May 2017; pp. 183–188. [CrossRef]

128. Malekzadeh, M.; Clegg, R.G.; Cavallaro, A.; Haddadi, H. Protecting Sensory Data Against Sensitive Inferences. In Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems (W-P2DS'18), Porto, Portugal, 23–26 April 2018; pp. 2:1–2:6. [CrossRef]

129. Malekzadeh, M.; Clegg, R.G.; Cavallaro, A.; Haddadi, H. Mobile Sensor Data Anonymization. In Proceedings of the International Conference on Internet of Things Design and Implementation, (IoTDI'19), Montreal, QC, Canada, 15–18 April 2019; pp. 49–58. [CrossRef]

130. Gao, X.; Luo, H.; Wang, Q.; Zhao, F.; Ye, L.; Zhang, Y. A Human Activity Recognition Algorithm Based on Stacking Denoising Autoencoder and LightGBM. *Sensors* **2019**, *19*, 947. [CrossRef]

131. Bai, L.; Yeung, C.; Efstratiou, C.; Chikomo, M. Motion2Vector: Unsupervised learning in human activity recognition using wrist-sensing data. In Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, London, UK, 9–13 September 2019; pp. 537–542.

132. Saeed, A.; Ozcelebi, T.; Lukkien, J.J. Synthesizing and Reconstructing Missing Sensory Modalities in Behavioral Context Recognition. *Sensors* **2018**, *18*, 2967. [CrossRef]

133. Wang, Y.; Wang, C.; Wang, Z.; Wang, X.; Li, Y. Hand gesture recognition using sparse autoencoder-based deep neural network based on electromyography measurements. In *Nano-, Bio-, Info-Tech Sensors, and 3D Systems II*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10597, p. 105971D.

134. Balabka, D. Semi-supervised learning for human activity recognition using adversarial autoencoders. In Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, London, UK, 9–13 September 2019; pp. 685–688.

135. De Andrade, F.H.C.; Pereira, F.G.; Resende, C.Z.; Cavalieri, D.C. Improving sEMG-Based Hand Gesture Recognition Using Maximal Overlap Discrete Wavelet Transform and an Autoencoder Neural Network. In *XXVI Brazilian Congress on Biomedical Engineering*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 271–279.

136. Chung, E.A.; Benalcázar, M.E. Real-Time Hand Gesture Recognition Model Using Deep Learning Techniques and EMG Signals. In Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2–6 September 2019; pp. 1–5.

137. Munoz-Organero, M.; Ruiz-Blazquez, R. Time-elastic generative model for acceleration time series in human activity recognition. *Sensors* **2017**, *17*, 319. [CrossRef]

138. Centeno, M.P.; van Moorsel, A.; Castruccio, S. Smartphone Continuous Authentication Using Deep Learning Autoencoders. In Proceedings of the 2017 15th Annual Conference on Privacy, Security and Trust (PST), Calgary, AB, Canada, 28–30 August 2017; pp. 147–1478. [CrossRef]

139. Vu, C.C.; Kim, J. Human motion recognition by textile sensors based on machine learning algorithms. *Sensors* **2018**, *18*, 3109. [CrossRef]

140. Chikhaoui, B.; Gouineau, F. Towards automatic feature extraction for activity recognition from wearable sensors: A deep learning approach. In Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, 18–21 November 2017; pp. 693–702.

141. Wang, L. Recognition of human activities using continuous autoencoders with wearable sensors. *Sensors* **2016**, *16*, 189. [CrossRef] [PubMed]

142. Jun, K.; Choi, S. Unsupervised End-to-End Deep Model for Newborn and Infant Activity Recognition. *Sensors* **2020**, *20*, 6467. [CrossRef] [PubMed]

143. Akbari, A.; Jafari, R. Transferring activity recognition models for new wearable sensors with deep generative domain adaptation. In Proceedings of the 18th International Conference on Information Processing in Sensor Networks, Montreal, QC, Canada, 16–18 April 2019; pp. 85–96.

144. Khan, M.A.A.H.; Roy, N. Untran: Recognizing unseen activities with unlabeled data using transfer learning. In Proceedings of the 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 17–20 April 2018; pp. 37–47.

145. Akbari, A.; Jafari, R. An autoencoder-based approach for recognizing null class in activities of daily living in-the-wild via wearable motion sensors. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3392–3396.

146. Prabono, A.G.; Yahya, B.N.; Lee, S.L. Atypical sample regularizer autoencoder for cross-domain human activity recognition. *Inf. Syst. Front.* **2021**, *23*, 71–80. [CrossRef]

147. Garcia, K.D.; de Sá, C.R.; Poel, M.; Carvalho, T.; Mendes-Moreira, J.; Cardoso, J.M.; de Carvalho, A.C.; Kok, J.N. An ensemble of autonomous auto-encoders for human activity recognition. *Neurocomputing* **2021**, *439*, 271–280. [CrossRef]

148. Valarezo, A.E.; Rivera, L.P.; Park, H.; Park, N.; Kim, T.S. Human activities recognition with a single wrist IMU via a Variational Autoencoder and android deep recurrent neural nets. *Comput. Sci. Inf. Syst.* **2020**, *17*, 581–597. [CrossRef]

149. Sigcha, L.; Costa, N.; Pavón, I.; Costa, S.; Arezes, P.; López, J.M.; De Arcas, G. Deep learning approaches for detecting freezing of gait in Parkinson's disease patients through on-body acceleration sensors. *Sensors* **2020**, *20*, 1895. [CrossRef]

150. Vavoulas, G.; Chatzaki, C.; Malliotakis, T.; Pediaditis, M.; Tsiknakis, M. The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones. In Proceedings of the ICT4AgeingWell, Rome, Italy, 21–22 April 2016; pp. 143–151.

151. Abu Alsheikh, M.; Selim, A.; Niyato, D.; Doyle, L.; Lin, S.; Tan, H.P. Deep Activity Recognition Models with Triaxial Accelerometers. *arXiv* **2015**, arXiv:1511.04664.

152. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. [CrossRef]

153. Gu, W.; Wang, G.; Zhang, Z.; Mao, Y.; Xie, X.; He, Y. High Accuracy Drug-Target Protein Interaction Prediction Method based on DBN. In Proceedings of the 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 28–30 July 2020; pp. 58–62.

154. Lefevre, F. A DBN-based multi-level stochastic spoken language understanding system. In Proceedings of the 2006 IEEE Spoken Language Technology Workshop, Palm Beach, Aruba, 10–13 December 2006; pp. 78–81.

155. Zhang, C.; He, Y.; Yuan, L.; Xiang, S. Analog circuit incipient fault diagnosis method using DBN based features extraction. *IEEE Access* **2018**, *6*, 23053–23064. [CrossRef]

156. Zhang, L.; Wu, X.; Luo, D. Real-Time Activity Recognition on Smartphones Using Deep Neural Networks. In Proceedings of the 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, China, 10–14 August 2015; pp. 1236–1242. [CrossRef]

157. Zhang, L.; Wu, X.; Luo, D. Recognizing Human Activities from Raw Accelerometer Data Using Deep Neural Networks. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 865–870. [CrossRef]

158. Radu, V.; Lane, N.D.; Bhattacharya, S.; Mascolo, C.; Marina, M.K.; Kawsar, F. Towards Multimodal Deep Learning for Activity Recognition on Mobile Devices. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, (UbiComp'16), Heidelberg, Germany, 12–16 September 2016; pp. 185–188. [CrossRef]

159. Mahmoodzadeh, A. Human Activity Recognition based on Deep Belief Network Classifier and Combination of Local and Global Features. *J. Inf. Syst. Telecommun.* **2021**, *9*, 33. [CrossRef]

160. Gao, Y.; Zhang, N.; Wang, H.; Ding, X.; Ye, X.; Chen, G.; Cao, Y. iHear Food: Eating Detection Using Commodity Bluetooth Headsets. In Proceedings of the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Washington, DC, USA, 27–29 June 2016; pp. 163–172. [CrossRef]

161. Hinton, G.E. A practical guide to training restricted Boltzmann machines. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 599–619.

162. Sani, S.; Massie, S.; Wiratunga, N.; Cooper, K. Learning deep and shallow features for human activity recognition. In *International Conference on Knowledge Science, Engineering and Management*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 469–482.

163. Matsui, S.; Inoue, N.; Akagi, Y.; Nagino, G.; Shinoda, K. User adaptation of convolutional neural network for human activity recognition. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, August 28–2 September 2017; pp. 753–757.

164. Akbari, A.; Jafari, R. Transition-Aware Detection of Modes of Locomotion and Transportation Through Hierarchical Segmentation. *IEEE Sens. J.* **2020**, *21*, 3301–3313. [CrossRef]

165. Feng, Y.; Chen, W.; Wang, Q. A strain gauge based locomotion mode recognition method using convolutional neural network. *Adv. Robot.* **2019**, *33*, 254–263. [CrossRef]

166. Ronao, C.A.; Cho, S.B. Deep convolutional neural networks for human activity recognition with smartphone sensors. In *International Conference on Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 46–53.

167. Huang, W.; Zhang, L.; Gao, W.; Min, F.; He, J. Shallow Convolutional Neural Networks for Human Activity Recognition Using Wearable Sensors. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]

168. Gao, W.; Zhang, L.; Huang, W.; Min, F.; He, J.; Song, A. Deep neural networks for sensor-based human activity recognition using selective kernel convolution. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [CrossRef]

169. Dong, M.; Han, J.; He, Y.; Jing, X. HAR-Net: Fusing Deep Representation and Hand-Crafted Features for Human Activity Recognition. In *Signal and Information Processing, Networking and Computers*; Sun, S., Fu, M., Xu, L., Eds.; Springer: Singapore, 2019; pp. 32–40.

170. Ravi, D.; Wong, C.; Lo, B.; Yang, G.Z. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In Proceedings of the 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), San Francisco, CA, USA, 14–17 June 2016; pp. 71–76.

171. Gholamrezaii, M.; AlModarresi, S.M.T. A time-efficient convolutional neural network model in human activity recognition. *Multimed. Tools Appl.* **2021**, *80*, 19361–19376. [CrossRef]

172. Xu, Y.; Qiu, T.T. Human Activity Recognition and Embedded Application Based on Convolutional Neural Network. *J. Artif. Intell. Technol.* **2021**, *1*, 51–60. [CrossRef]

173. Uddin, M.Z.; Hassan, M.M. Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. *IEEE Sens. J.* **2018**, *19*, 8413–8419. [CrossRef]

174. Ha, S.; Choi, S. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 381–388.

175. Jordao, A.; Torres, L.A.B.; Schwartz, W.R. Novel approaches to human activity recognition based on accelerometer data. *Signal Image Video Process.* **2018**, *12*, 1387–1394. [CrossRef]

176. Grzeszick, R.; Lenk, J.M.; Rueda, F.M.; Fink, G.A.; Feldhorst, S.; ten Hompel, M. Deep neural network based human activity recognition for the order picking process. In Proceedings of the 4th International Workshop on Sensor-Based Activity Recognition and Interaction, Rostock, Germany, 21–22 September 2017; pp. 1–6.

177. Gao, W.; Zhang, L.; Teng, Q.; He, J.; Wu, H. DanHAR: Dual Attention Network for multimodal human activity recognition using wearable sensors. *Appl. Soft Comput.* **2021**, *111*, 107728. [CrossRef]

178. Murakami, K.; Taguchi, H. Gesture Recognition Using Recurrent Neural Networks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (CHI'91), New Orleans, LA, USA, 27 April–2 May 1991; pp. 237–242. [CrossRef]

179. Vamplew, P.; Adams, A. Recognition and anticipation of hand motions using a recurrent neural network. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 6, pp. 2904. [CrossRef]

180. Long, Y.; Jung, E.M.; Kung, J.; Mukhopadhyay, S. Reram crossbar based recurrent neural network for human activity detection. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 939–946.

181. Inoue, M.; Inoue, S.; Nishida, T. Deep recurrent neural network for mobile human activity recognition with high throughput. *Artif. Life Robot.* **2018**, *23*, 173–185. [CrossRef]

182. Chen, W.; Wang, S.; Zhang, X.; Yao, L.; Yue, L.; Qian, B.; Li, X. EEG-based motion intention recognition via multi-task RNNs. In Proceedings of the 2018 SIAM International Conference on Data Mining, San Diego, CA, USA, 3–5 May 2018; pp. 279–287.

183. Carfi, A.; Motolese, C.; Bruno, B.; Mastrogiovanni, F. Online human gesture recognition using recurrent neural networks and wearable sensors. In Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Nanjing, China, 27–31 August 2018; pp. 188–195.

184. Tamamori, A.; Hayashi, T.; Toda, T.; Takeda, K. An investigation of recurrent neural network for daily activity recognition using multi-modal signals. In Proceedings of the 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, 12–15 December 2017; pp. 1334–1340.

185. Tamamori, A.; Hayashi, T.; Toda, T.; Takeda, K. Daily activity recognition based on recurrent neural network using multi-modal signals. *APSIPA Trans. Signal Inf. Process.* **2018**, 7, e21. [CrossRef]

186. Uddin, M.Z.; Hassan, M.M.; Alsanad, A.; Savaglio, C. A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. *Inf. Fusion* **2020**, 55, 105–115. [CrossRef]

187. Alessandrini, M.; Biagetti, G.; Crippa, P.; Falaschetti, L.; Turchetti, C. Recurrent Neural Network for Human Activity Recognition in Embedded Systems Using PPG and Accelerometer Data. *Electronics* **2021**, 10, 1715. [CrossRef]

188. Zheng, L.; Li, S.; Zhu, C.; Gao, Y. Application of IndRNN for human activity recognition: The Sussex-Huawei locomotion-transportation challenge. In Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, London, UK, 9–13 September 2019; pp. 869–872. [CrossRef]

189. Bailador, G.; Roggen, D.; Tröster, G.; Triviño, G. Real time gesture recognition using continuous time recurrent neural networks. In *BodyNets*; Citeseer: Princeton, NJ, USA, 2007; p. 15.

190. Wang, X.; Liao, W.; Guo, Y.; Yu, L.; Wang, Q.; Pan, M.; Li, P. Perrnn: Personalized recurrent neural networks for acceleration-based human activity recognition. In Proceedings of the ICC 2019–2019 IEEE International Conference on Communications (ICC) Shanghai, China, 20–24 May 2019; pp. 1–6. [CrossRef]

191. Khatiwada, P.; Subedi, M.; Chatterjee, A.; Gerdes, M.W. Automated Human Activity Recognition by Colliding Bodies Optimization-based Optimal Feature Selection with Recurrent Neural Network. *arXiv* **2020**, arXiv:2010.03324.

192. Lv, M.; Xu, W.; Chen, T. A hybrid deep convolutional and recurrent neural network for complex activity recognition using multimodal sensors. *Neurocomputing* **2019**, 362, 33–40. [CrossRef]

193. Ketykó, I.; Kovács, F.; Varga, K.Z. Domain adaptation for semg-based gesture recognition with recurrent neural networks. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–7. [CrossRef]

194. Bengio, Y.; Frasconi, P.; Schmidhuber, J.; Elvezia, C. Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term D Ependencies* Sepp Hochreiter Fakult at f ur Informatik. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.24.7321&rep=rep1&type=pdf (accessed on 5 November 2021).

195. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, 9, 1735–1780. [CrossRef]

196. Chung, J.; Gülçehre, Ç.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.

197. Czuszyński, K.; Rumiński, J.; Kwaśniewska, A. Gesture recognition with the linear optical sensor and recurrent neural networks. *IEEE Sens. J.* **2018**, 18, 5429–5438. [CrossRef]

198. Zhu, Y.; Luo, H.; Chen, R.; Zhao, F.; Su, L. DenseNetX and GRU for the Sussex-Huawei locomotion-transportation recognition challenge. In Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, Virtual, 12–17 September 2020; pp. 373–377. [CrossRef]

199. Okai, J.; Paraschiakos, S.; Beekman, M.; Knobbe, A.; de Sá, C.R. Building robust models for human activity recognition from raw accelerometers data using gated recurrent units and long short term memory neural networks. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2486–2491. [CrossRef]

200. Hu, Y.; Zhang, X.Q.; Xu, L.; He, F.X.; Tian, Z.; She, W.; Liu, W. Harmonic Loss Function for Sensor-Based Human Activity Recognition Based on LSTM Recurrent Neural Networks. *IEEE Access* **2020**, 8, 135617–135627. [CrossRef]

201. Jangir, M.K.; Singh, K. HARGRURNN: Human activity recognition using inertial body sensor gated recurrent units recurrent neural network. *J. Discret. Math. Sci. Cryptogr.* **2019**, 22, 1577–1587. [CrossRef]

202. Murad, A.; Pyun, J.Y. Deep Recurrent Neural Networks for Human Activity Recognition. *Sensors* **2017**, 17, 2556. [CrossRef] [PubMed]

203. Rivera, P.; Valarezo, E.; Choi, M.T.; Kim, T.S. Recognition of human hand activities based on a single wrist imu using recurrent neural networks. *Int. J. Pharm. Med. Biol. Sci.* **2017**, 6, 114–118. [CrossRef]

204. Hammerla, N.Y.; Halloran, S.; Plötz, T. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv* **2016**, arXiv:1604.08880.

205. Li, C.; Xie, C.; Zhang, B.; Chen, C.; Han, J. Deep Fisher discriminant learning for mobile hand gesture recognition. *Pattern Recognit.* **2018**, 77, 276–288. [CrossRef]

206. Tao, D.; Wen, Y.; Hong, R. Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition. *IEEE Internet Things J.* **2016**, 3, 1124–1134. [CrossRef]

207. Zebin, T.; Sperrin, M.; Peek, N.; Casson, A.J. Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 1–4. [CrossRef]

208. Ordóñez, F.J.; Roggen, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [CrossRef]

209. Barut, O.; Zhou, L.; Luo, Y. Multitask LSTM Model for Human Activity Recognition and Intensity Estimation Using Wearable Sensor Data. *IEEE Internet Things J.* **2020**, *7*, 8760–8768. [CrossRef]

210. Qin, Y.; Luo, H.; Zhao, F.; Wang, C.; Wang, J.; Zhang, Y. Toward Transportation Mode Recognition Using Deep Convolutional and Long Short-Term Memory Recurrent Neural Networks. *IEEE Access* **2019**, *7*, 142353–142367. [CrossRef]

211. Xia, K.; Huang, J.; Wang, H. LSTM-CNN architecture for human activity recognition. *IEEE Access* **2020**, *8*, 56855–56866. [CrossRef]

212. Wu, Y.; Zheng, B.; Zhao, Y. Dynamic gesture recognition based on LSTM-CNN. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 2446–2450. [CrossRef]

213. Friedrich, B.; Lübbe, C.; Hein, A. Combining LSTM and CNN for mode of transportation classification from smartphone sensors. In Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, Virtual Event, 12–17 September 2020; pp. 305–310. [CrossRef]

214. Senyurek, V.Y.; Imtiaz, M.H.; Belsare, P.; Tiffany, S.; Sazonov, E. A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors. *Biomed. Eng. Lett.* **2020**, *10*, 195–203. [CrossRef]

215. Mutegeki, R.; Han, D.S. A CNN-LSTM approach to human activity recognition. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 19–21 February 2020; pp. 362–366. [CrossRef]

216. Ahmad, W.; Kazmi, B.M.; Ali, H. Human activity recognition using multi-head CNN followed by LSTM. In Proceedings of the 2019 15th International Conference on Emerging Technologies (ICET), Peshawar, Pakistan, 2–3 December 2019; pp. 1–6. [CrossRef]

217. Deep, S.; Zheng, X. Hybrid model featuring CNN and LSTM architecture for human activity recognition on smartphone sensor data. In Proceedings of the 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Gold Coast, Australia, 5–7 December 2019; pp. 259–264. [CrossRef]

218. Gao, J.; Gu, P.; Ren, Q.; Zhang, J.; Song, X. Abnormal gait recognition algorithm based on LSTM-CNN fusion network. *IEEE Access* **2019**, *7*, 163180–163190. [CrossRef]

219. Saha, S.S.; Sandha, S.S.; Srivastava, M. Deep Convolutional Bidirectional LSTM for Complex Activity Recognition with Missing Data. In *Human Activity Recognition Challenge*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 39–53.

220. Mekruksavanich, S.; Jitpattanakul, A.; Thongkum, P. Placement Effect of Motion Sensors for Human Activity Recognition using LSTM Network. In Proceedings of the 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, Cha-am, Thailand, 3–6 March 2021; pp. 273–276. [CrossRef]

221. Chéron, G.; Leurs, F.; Bengoetxea, A.; Draye, J.; Destrée, M.; Dan, B. A dynamic recurrent neural network for multiple muscles electromyographic mapping to elevation angles of the lower limb in human locomotion. *J. Neurosci. Methods* **2003**, *129*, 95–104. [CrossRef]

222. Gupta, R.; Dhindsa, I.S.; Agarwal, R. Continuous angular position estimation of human ankle during unconstrained locomotion. *Biomed. Signal Process. Control.* **2020**, *60*, 101968. [CrossRef]

223. Hioki, M.; Kawasaki, H. Estimation of finger joint angles from sEMG using a recurrent neural network with time-delayed input vectors. In Proceedings of the 2009 IEEE International Conference on Rehabilitation Robotics, Kyoto, Japan, 23–26 June 2009; pp. 289–294. [CrossRef]

224. Bu, N.; Fukuda, O.; Tsuji, T. EMG-based motion discrimination using a novel recurrent neural network. *J. Intell. Inf. Syst.* **2003**, *21*, 113–126. [CrossRef]

225. Cheron, G.; Cebolla, A.M.; Bengoetxea, A.; Leurs, F.; Dan, B. Recognition of the physiological actions of the triphasic EMG pattern by a dynamic recurrent neural network. *Neurosci. Lett.* **2007**, *414*, 192–196. [CrossRef]

226. Zeng, M.; Gao, H.; Yu, T.; Mengshoel, O.J.; Langseth, H.; Lane, I.; Liu, X. Understanding and improving recurrent networks for human activity recognition by continuous attention. In Proceedings of the 2018 ACM International Symposium on Wearable Computers, Singapore, 8–12 October 2018; pp. 56–63. [CrossRef]

227. Xu, C.; Chai, D.; He, J.; Zhang, X.; Duan, S. InnoHAR: A Deep Neural Network for Complex Human Activity Recognition. *IEEE Access* **2019**, *7*, 9893–9902. [CrossRef]

228. Qian, H.; Pan, S.J.; Da, B.; Miao, C. A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition. *IJCAI* **2019**, *2019*, 5614–5620.

229. Kung-Hsiang (Steeve), H. Introduction to Various Reinforcement Learning Algorithms. Part I (Q-Learning, SARSA, DQN, DDPG). 2018. Available online: https://towardsdatascience.com/introduction-to-various-reinforcement-learning-algorithms-i-q-learning-sarsa-dqn-ddpg-72a5e0cb6287 (accessed on 13 July 2012).

230. Seok, W.; Kim, Y.; Park, C. Pattern recognition of human arm movement using deep reinforcement learning. In Proceedings of the 2018 International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, 10–12 January 2018; pp. 917–919. [CrossRef]

231. Zheng, J.; Cao, H.; Chen, D.; Ansari, R.; Chu, K.C.; Huang, M.C. Designing deep reinforcement learning systems for musculoskeletal modeling and locomotion analysis using wearable sensor feedback. *IEEE Sens. J.* **2020**, *20*, 9274–9282. [CrossRef]

232. Bhat, G.; Deb, R.; Chaurasia, V.V.; Shill, H.; Ogras, U.Y. Online human activity recognition using low-power wearable devices. In Proceedings of the 2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Diego, CA, USA, 5–8 November 2018; pp. 1–8. [CrossRef]

233. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, 2672–2680. Available online: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b1 22f61f8f06494c97b1afccf3-Paper.pdf (accessed on 5 November 2021).

234. Farnia, F.; Ozdaglar, A. Gans may have no nash equilibria. *arXiv* **2020**, arXiv:2002.09124.

235. Springenberg, J.T. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv* **2015**, arXiv:1511.06390.

236. Odena, A. Semi-supervised learning with generative adversarial networks. *arXiv* **2016**, arXiv:1606.01583.

237. Shi, J.; Zuo, D.; Zhang, Z. A GAN-based data augmentation method for human activity recognition via the caching ability. *Internet Technol. Lett.* **2021**, *4*, e257. [CrossRef]

238. Wang, J.; Chen, Y.; Gu, Y.; Xiao, Y.; Pan, H. SensoryGANs: An Effective Generative Adversarial Framework for Sensor-based Human Activity Recognition. In Proceedings of the 2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8. [CrossRef]

239. Kawaguchi, N.; Yang, Y.; Yang, T.; Ogawa, N.; Iwasaki, Y.; Kaji, K.; Terada, T.; Murao, K.; Inoue, S.; Kawahara, Y.; et al. HASC2011corpus: Towards the Common Ground of Human Activity Recognition. In Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp'11), Beijing, China, 17–21 September 2011; pp. 571–572. [CrossRef]

240. Alharbi, F.; Ouarbya, L.; Ward, J.A. Synthetic sensor data for human activity recognition. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–9. [CrossRef]

241. Chan, M.H.; Noor, M.H.M. A unified generative model using generative adversarial network for activity recognition. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *12*, 8119–8128. [CrossRef]

242. Li, X.; Luo, J.; Younes, R. ActivityGAN: Generative adversarial networks for data augmentation in sensor-based human activity recognition. In Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, Virtual Event Mexico, 12–17 September 2020; pp. 249–254.

243. Shi, X.; Li, Y.; Zhou, F.; Liu, L. Human activity recognition based on deep learning method. In Proceedings of the 2018 International Conference on Radar (RADAR), Brisbane, QLD, Australia, 27–31 August 2018; pp. 1–5. [CrossRef]

244. Soleimani, E.; Nazerfard, E. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing* **2021**, *426*, 26–34. [CrossRef]

245. Abedin, A.; Rezatofighi, H.; Ranasinghe, D.C. Guided-GAN: Adversarial Representation Learning for Activity Recognition with Wearables. *arXiv* **2021**, arXiv:2110.05732.

246. Sanabria, A.R.; Zambonelli, F.; Dobson, S.; Ye, J. ContrasGAN: Unsupervised domain adaptation in Human Activity Recognition via adversarial and contrastive learning. *Pervasive Mob. Comput.* **2021**, *78*, 101477. [CrossRef]

247. Challa, S.K.; Kumar, A.; Semwal, V.B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *Vis. Comput.* **2021**, 1–15. [CrossRef]

248. Dua, N.; Singh, S.N.; Semwal, V.B. Multi-input CNN-GRU based human activity recognition using wearable sensors. *Computing* **2021**, *103*, 1461–1478. [CrossRef]

249. Zhang, X.; Yao, L.; Huang, C.; Wang, S.; Tan, M.; Long, G.; Wang, C. Multi-modality Sensor Data Classification with Selective Attention. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, (IJCAI-18). International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 13–19 July 2018; pp. 3111–3117. [CrossRef]

250. Yao, L.; Sheng, Q.Z.; Li, X.; Gu, T.; Tan, M.; Wang, X.; Wang, S.; Ruan, W. Compressive representation for device-free activity recognition with passive RFID signal strength. *IEEE Trans. Mob. Comput.* **2017**, *17*, 293–306. [CrossRef]

251. Zhang, X.; Yao, L.; Wang, X.; Zhang, W.; Zhang, S.; Liu, Y. Know your mind: Adaptive cognitive activity recognition with reinforced CNN. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; pp. 896–905. [CrossRef]

252. Wan, S.; Qi, L.; Xu, X.; Tong, C.; Gu, Z. Deep learning models for real-time human activity recognition with smartphones. *Mob. Netw. Appl.* **2020**, *25*, 743–755. [CrossRef]

253. Zhao, Y.; Yang, R.; Chevalier, G.; Xu, X.; Zhang, Z. Deep residual bidir-LSTM for human activity recognition using wearable sensors. *Math. Probl. Eng.* **2018**, *2018*, 7316954. [CrossRef]

254. Ullah, M.; Ullah, H.; Khan, S.D.; Cheikh, F.A. Stacked lstm network for human activity recognition using smartphone data. In Proceedings of the 2019 8th European workshop on visual information processing (EUVIP), Roma, Italy, 28–31 October 2019; pp. 175–180.

255. Hernández, F.; Suárez, L.F.; Villamizar, J.; Altuve, M. Human activity recognition on smartphones using a bidirectional lstm network. In Proceedings of the 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Bucaramanga, Colombia, 24–26 April 2019; pp. 1–5.

256. Cheng, X.; Zhang, L.; Tang, Y.; Liu, Y.; Wu, H.; He, J. Real-time Human Activity Recognition Using Conditionally Parametrized Convolutions on Mobile and Wearable Devices. *arXiv* **2020**, arXiv:2006.03259.

257. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.

258. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein gans. *arXiv* **2017**, arXiv:1704.00028.

259. Che, T.; Li, Y.; Jacob, A.P.; Bengio, Y.; Li, W. Mode regularized generative adversarial networks. *arXiv* **2016**, arXiv:1612.02136.

260. Gao, Y.; Jin, Y.; Chauhan, J.; Choi, S.; Li, J.; Jin, Z. Voice In Ear: Spoofing-Resistant and Passphrase-Independent Body Sound Authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–25. [CrossRef]

261. Steven Eyobu, O.; Han, D.S. Feature Representation and Data Augmentation for Human Activity Classification Based on Wearable IMU Sensor Data Using a Deep LSTM Neural Network. *Sensors* **2018**, *18*, 2892. [CrossRef]

262. Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Data augmentation using synthetic data for time series classification with deep residual networks. In Proceedings of the International Workshop on Advanced Analytics and Learning on Temporal Data, ECML PKDD, Dublin, Ireland, 10–14 September 2018.

263. Ramponi, G.; Protopapas, P.; Brambilla, M.; Janssen, R. T-CGAN: Conditional Generative Adversarial Network for Data Augmentation in Noisy Time Series with Irregular Sampling. *arXiv* **2018**, arXiv:1811.08295.

264. Alzantot, M.; Chakraborty, S.; Srivastava, M. SenseGen: A deep learning architecture for synthetic sensor data generation. In Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kona, HI, USA, 13–17 March 2017. [CrossRef]

265. Kwon, H.; Tong, C.; Haresamudram, H.; Gao, Y.; Abowd, G.D.; Lane, N.D.; Plötz, T. IMUTube: Automatic Extraction of Virtual on-Body Accelerometry from Video for Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 1–29. [CrossRef]

266. Liu, Y.; Zhang, S.; Gowda, M. When Video Meets Inertial Sensors: Zero-Shot Domain Adaptation for Finger Motion Analytics with Inertial Sensors. In Proceedings of the International Conference on Internet-of-Things Design and Implementation (IoTDI'21), Charlottesvle, VA, USA, 18–21 May 2021; pp. 182–194. [CrossRef]

267. Zhou, Y.; Wang, Z.; Fang, C.; Bui, T.; Berg, T.L. Visual to sound: Generating natural sound for videos in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3550–3558.

268. Hossain, M.Z.; Sohel, F.; Shiratuddin, M.F.; Laga, H. A Comprehensive Survey of Deep Learning for Image Captioning. *ACM Comput. Surv.* **2019**, *51*, 1–36. [CrossRef]

269. Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; Lee, H. Generative Adversarial Text to Image Synthesis. In Proceedings of The 33rd International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; Volume 48, pp. 1060–1069.

270. Zhang, S.; Alshurafa, N. Deep Generative Cross-Modal on-Body Accelerometer Data Synthesis from Videos. In Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, (UbiComp-ISWC'20), Virtual, 12–17 September 2020; pp. 223–227. [CrossRef]

271. Rahman, M.; Ali, N.; Bari, R.; Saleheen, N.; al'Absi, M.; Ertin, E.; Kennedy, A.; Preston, K.L.; Kumar, S., mDebugger: Assessing and Diagnosing the Fidelity and Yield of Mobile Sensor Data. In *Mobile Health: Sensors, Analytic Methods, and Applications*; Rehg, J.M., Murphy, S.A., Kumar, S., Eds.; Springer: Cham, Switzerland, 2017; pp. 121–143. [CrossRef]

272. Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; Li, Y. BRITS: Bidirectional Recurrent Imputation for Time Series. In *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; pp. 6775–6785.

273. Luo, Y.; Cai, X.; Zhang, Y.; Xu, J. Multivariate Time Series Imputation with Generative Adversarial Networks. In *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; pp. 1596–1607.

274. Rolnick, D.; Veit, A.; Belongie, S.J.; Shavit, N. Deep Learning is Robust to Massive Label Noise. *arXiv* **2017**, arXiv:1705.10694.

275. Mothukuri, V.; Parizi, R.M.; Pouriyeh, S.; Huang, Y.; Dehghantanha, A.; Srivastava, G. A survey on security and privacy of federated learning. *Future Gener. Comput. Syst.* **2021**, *115*, 619–640. [CrossRef]

276. Briggs, C.; Fan, Z.; Andras, P. A review of privacy-preserving federated learning for the Internet-of-Things. *Fed. Learn. Syst.* **2021**, pp. 21–50.

277. Sozinov, K.; Vlassov, V.; Girdzijauskas, S. Human activity recognition using federated learning. In Proceedings of the 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), Melbourne, VIC, Australia, 11–13 December 2018; pp. 1103–1111. [CrossRef]

278. Li, C.; Niu, D.; Jiang, B.; Zuo, X.; Yang, J. Meta-HAR: Federated Representation Learning for Human Activity Recognition. In Proceedings of the Web Conference 2021 (WWW'21); Association for Computing Machinery: Ljubljana, Slovenia, 12–23 April 2021; pp. 912–922. [CrossRef]

279. Xiao, Z.; Xu, X.; Xing, H.; Song, F.; Wang, X.; Zhao, B. A federated learning system with enhanced feature extraction for human activity recognition. *Knowl. Based Syst.* **2021**, *229*, 107338. [CrossRef]

280. Tu, L.; Ouyang, X.; Zhou, J.; He, Y.; Xing, G. FedDL: Federated Learning via Dynamic Layer Sharing for Human Activity Recognition. In Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra Portugal, 15–17 November 2021; pp. 15–28. [CrossRef]

281. Bettini, C.; Civitarese, G.; Presotto, R. Personalized Semi-Supervised Federated Learning for Human Activity Recognition. *arXiv* **2021**, arXiv:2104.08094.

282. Gudur, G.K.; Perepu, S.K. Resource-constrained federated learning with heterogeneous labels and models for human activity recognition. In Proceedings of the Deep Learning for Human Activity Recognition: Second International Workshop, DL-HAR 2020, Kyoto, Japan, 8 January 2021; Springer: Berlin/Heidelberg, Germany, 2021, Volume 1370, p. 57.

283. Bdiwi, R.; de Runz, C.; Faiz, S.; Cherif, A.A. Towards a New Ubiquitous Learning Environment Based on Blockchain Technology. In Proceedings of the 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), Timisoara, Romania, 3–7 July 2017; pp. 101–102. [CrossRef]

284. Bdiwi, R.; De Runz, C.; Faiz, S.; Cherif, A.A. A blockchain based decentralized platform for ubiquitous learning environment. In Proceedings of the 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, 9–13 July 2018; pp. 90–92. [CrossRef]

285. Shrestha, A.K.; Vassileva, J.; Deters, R. A Blockchain Platform for User Data Sharing Ensuring User Control and Incentives. *Front. Blockchain* **2020**, *3*, 48. [CrossRef]

286. Chen, Z.; Fiandrino, C.; Kantarci, B. On blockchain integration into mobile crowdsensing via smart embedded devices: A comprehensive survey. *J. Syst. Archit.* **2021**, *115*, 102011. [CrossRef]

287. Nguyen, D.C.; Ding, M.; Pham, Q.V.; Pathirana, P.N.; Le, L.B.; Seneviratne, A.; Li, J.; Niyato, D.; Poor, H.V. Federated learning meets blockchain in edge computing: Opportunities and challenges. *IEEE Internet Things J.* **2021**, *8*, 12806–12825. [CrossRef]

288. Zhang, Y.C.; Zhang, S.; Liu, M.; Daly, E.; Battalio, S.; Kumar, S.; Spring, B.; Rehg, J.M.; Alshurafa, N. SyncWISE: Window Induced Shift Estimation for Synchronization of Video and Accelerometry from Wearable Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 1–26. [CrossRef]

289. Fridman, L.; Brown, D.E.; Angell, W.; Abdic, I.; Reimer, B.; Noh, H.Y. Automated Synchronization of Driving Data Using Vibration and Steering Events. *Pattern Recognit. Lett.* **2015**, *75*, 9–15. [CrossRef]

290. Zeng, M.; Yu, T.; Wang, X.; Nguyen, L.T.; Mengshoel, O.J.; Lane, I. Semi-supervised convolutional neural networks for human activity recognition. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 522–529. [CrossRef]

291. Chen, K.; Yao, L.; Zhang, D.; Chang, X.; Long, G.; Wang, S. Distributionally Robust Semi-Supervised Learning for People-Centric Sensing. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 3321–3328. [CrossRef]

292. Gudur, G.K.; Sundaramoorthy, P.; Umaashankar, V. ActiveHARNet: Towards On-Device Deep Bayesian Active Learning for Human Activity Recognition. In Proceedings of the The 3rd International Workshop on Deep Learning for Mobile Systems and Applications, (EMDL'19), Seoul, Korea, 21 June 2019; pp. 7–12. [CrossRef]

293. Rizve, M.N.; Duarte, K.; Rawat, Y.S.; Shah, M. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv* **2021**, arXiv:2101.06329.

294. Alharbi, R.; Vafaie, N.; Liu, K.; Moran, K.; Ledford, G.; Pfammatter, A.; Spring, B.; Alshurafa, N. Investigating barriers and facilitators to wearable adherence in fine-grained eating detection. In Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kona, HI, USA, 13–17 March 2017; pp. 407–412. [CrossRef]

295. Nakamura, K.; Yeung, S.; Alahi, A.; Fei-Fei, L. Jointly Learning Energy Expenditures and Activities Using Egocentric Multimodal Signals. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6817–6826. [CrossRef]

296. Plötz, T.; Guan, Y. Deep Learning for Human Activity Recognition in Mobile Computing. *Computer* **2018**, *51*, 50–59. [CrossRef]

297. Qian, H.; Pan, S.J.; Miao, C. Weakly-supervised sensor-based activity segmentation and recognition via learning from distributions. *Artif. Intell.* **2021**, *292*, 103429. [CrossRef]

298. Kyritsis, K.; Diou, C.; Delopoulos, A. Modeling Wrist Micromovements to Measure In-Meal Eating Behavior from Inertial Sensor Data. *IEEE J. Biomed. Health Inform.* **2019**. [CrossRef]

299. Liu, C.; Zhang, L.; Liu, Z.; Liu, K.; Li, X.; Liu, Y. Lasagna: Towards Deep Hierarchical Understanding and Searching over Mobile Sensing Data. In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, (MobiCom'16), New York, NY, USA, 3–7 October 2016; pp. 334–347. [CrossRef]

300. Peng, L.; Chen, L.; Ye, Z.; Zhang, Y. AROMA: A Deep Multi-Task Learning Based Simple and Complex Human Activity Recognition Method Using Wearable Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 74:1–74:16. [CrossRef]

301. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Philip, S.Y. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [CrossRef]

302. Qin, Z.; Zhang, Y.; Meng, S.; Qin, Z.; Choo, K.K.R. Imaging and fusing time series for wearable sensor-based human activity recognition. *Inf. Fusion* **2020**, *53*, 80–87. [CrossRef]

303. Abdel-Basset, M.; Hawash, H.; Chang, V.; Chakrabortty, R.K.; Ryan, M. Deep learning for Heterogeneous Human Activity Recognition in Complex IoT Applications. *IEEE Internet Things J.* **2020**. [CrossRef]

304. Siirtola, P.; Röning, J. Incremental Learning to Personalize Human Activity Recognition Models: The Importance of Human AI Collaboration. *Sensors* **2019**, *19*, 5151. [CrossRef]

305. Qian, H.; Pan, S.J.; Miao, C.; Qian, H.; Pan, S.; Miao, C. Latent Independent Excitation for Generalizable Sensor-based Cross-Person Activity Recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February, 2021; Volume 35, pp. 11921–11929.

306. Arjovsky, M.; Bottou, L.; Gulrajani, I.; Lopez-Paz, D. Invariant Risk Minimization. *arXiv* **2020**, arXiv:1907.02893.

307. Konečnỳ, J.; McMahan, B.; Ramage, D. Federated optimization: Distributed optimization beyond the datacenter. *arXiv* **2015**, arXiv:1511.03575.

308. Qiu, S.; Zhao, H.; Jiang, N.; Wang, Z.; Liu, L.; An, Y.; Zhao, H.; Miao, X.; Liu, R.; Fortino, G. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion* **2022**, *80*, 241–265. [CrossRef]

309. Ahad, M.A.R.; Antar, A.D.; Ahmed, M. Sensor-based human activity recognition: Challenges ahead. In *IoT Sensor-Based Activity Recognition*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 175–189.

310. Abedin, A.; Ehsanpour, M.; Shi, Q.; Rezatofighi, H.; Ranasinghe, D.C. Attend and Discriminate: Beyond the State-of-the-Art for Human Activity Recognition Using Wearable Sensors. *Proc. Acm Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–22. [CrossRef]

311. Huynh-The, T.; Hua, C.H.; Tu, N.A.; Kim, D.S. Physical Activity Recognition with Statistical-Deep Fusion Model Using Multiple Sensory Data for Smart Health. *IEEE Internet Things J.* **2021**, *8*, 1533–1543. [CrossRef]

312. Hanif, M.; Akram, T.; Shahzad, A.; Khan, M.; Tariq, U.; Choi, J.; Nam, Y.; Zulfiqar, Z. Smart Devices Based Multisensory Approach for Complex Human Activity Recognition. *Comput. Mater. Contin.* **2022**, *70*, 3221–3234. [CrossRef]

313. Pires, I.M.; Pombo, N.; Garcia, N.M.; Flórez-Revuelta, F. Multi-Sensor Mobile Platform for the Recognition of Activities of Daily Living and Their Environments Based on Artificial Neural Networks. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, (IJCAI-18). International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 13–19 July 2018; pp. 5850–5852. [CrossRef]

314. Sena, J.; Barreto, J.; Caetano, C.; Cramer, G.; Schwartz, W.R. Human activity recognition based on smartphone and wearable sensors using multiscale DCNN ensemble. *Neurocomputing* **2021**, *444*, 226–243. [CrossRef]

315. Xia, S.; Chandrasekaran, R.; Liu, Y.; Yang, C.; Rosing, T.S.; Jiang, X. A Drone-Based System for Intelligent and Autonomous Homes. In Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems (SenSys'21), Coimbra, Portugal, 15–17 November 2021; pp. 349–350. [CrossRef]

316. Lane, N.D.; Bhattacharya, S.; Georgiev, P.; Forlivesi, C.; Jiao, L.; Qendro, L.; Kawsar, F. DeepX: A Software Accelerator for Low-Power Deep Learning Inference on Mobile Devices. In Proceedings of the 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Vienna, Austria, 11–14 April 2016; pp. 1–12. [CrossRef]

317. Lane, N.D.; Georgiev, P.; Qendro, L. DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments Using Deep Learning. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15), Osaka, Japan, 7–11 September 2015; pp. 283–294. [CrossRef]

318. Cao, Q.; Balasubramanian, N.; Balasubramanian, A. MobiRNN: Efficient Recurrent Neural Network Execution on Mobile GPU. In Proceedings of the 1st International Workshop on Deep Learning for Mobile Systems and Applications (EMDL'17), Niagara Falls, NY, USA, 23 June 2017; pp. 1–6. [CrossRef]

319. Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; Abdelzaher, T. DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing. In Proceedings of the 26th International Conference on World Wide Web; International World Wide Web Conferences Steering Committee: Republic and Canton of Geneva, Switzerland (WWW'17), Perth, Australia, 3–7 April 2017; pp. 351–360. [CrossRef]

320. Bhattacharya, S.; Lane, N.D. Sparsification and Separation of Deep Learning Layers for Constrained Resource Inference on Wearables. In Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM, (SenSys'16), Stanford, CA, USA, 14–16 November 2016; pp. 176–189. [CrossRef]

321. Edel, M.; Köppe, E. Binarized-BLSTM-RNN based Human Activity Recognition. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcala de Henares, Spain, 4–7 October 2016; pp. 1–7. [CrossRef]

322. Bhat, G.; Tuncel, Y.; An, S.; Lee, H.G.; Ogras, U.Y. An Ultra-Low Energy Human Activity Recognition Accelerator for Wearable Health Applications. *ACM Trans. Embed. Comput. Syst.* **2019**, *18*, 1–22. [CrossRef]

323. Wang, L.; Thiemjarus, S.; Lo, B.; Yang, G.Z. Toward a mixed-signal reconfigurable ASIC for real-time activity recognition. In Proceedings of the 2008 5th International Summer School and Symposium on Medical Devices and Biosensors, Hong Kong, China, 1–3 June 2008; pp. 227–230. [CrossRef]

324. Islam, B.; Nirjon, S. Zygarde: Time-Sensitive On-Device Deep Inference and Adaptation on Intermittently-Powered Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 1–29. [CrossRef]

325. Xia, S.; Nie, J.; Jiang, X. CSafe: An Intelligent Audio Wearable Platform for Improving Construction Worker Safety in Urban Environments. In Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021), (IPSN'21), Nashville, TN, USA, 18–21 May 2021; pp. 207–221. [CrossRef]

326. Xia, S.; de Godoy Peixoto, D.; Islam, B.; Islam, M.T.; Nirjon, S.; Kinget, P.R.; Jiang, X. Improving Pedestrian Safety in Cities Using Intelligent Wearable Systems. *IEEE Internet Things J.* **2019**, *6*, 7497–7514. [CrossRef]

327. de Godoy, D.; Islam, B.; Xia, S.; Islam, M.T.; Chandrasekaran, R.; Chen, Y.C.; Nirjon, S.; Kinget, P.R.; Jiang, X. PAWS: A Wearable Acoustic System for Pedestrian Safety. In Proceedings of the 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 17–20 April 2018; pp. 237–248. [CrossRef]

328. Nie, J.; Hu, Y.; Wang, Y.; Xia, S.; Jiang, X. SPIDERS: Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition. In Proceedings of the 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI), Sydney, NSW, Australia, 21–24 April 2020; pp. 27–39. [CrossRef]

329. Nie, J.; Liu, Y.; Hu, Y.; Wang, Y.; Xia, S.; Preindl, M.; Jiang, X. SPIDERS+: A light-weight, wireless, and low-cost glasses-based wearable platform for emotion sensing and bio-signal acquisition. *Pervasive Mob. Comput.* **2021**, *75*, 101424. [CrossRef]
330. Hu, Y.; Nie, J.; Wang, Y.; Xia, S.; Jiang, X. Demo Abstract: Wireless Glasses for Non-contact Facial Expression Monitoring. In Proceedings of the 2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Sydney, NSW, Australia, 21–24 April 2020; pp. 367–368. [CrossRef]
331. Chandrasekaran, R.; de Godoy, D.; Xia, S.; Islam, M.T.; Islam, B.; Nirjon, S.; Kinget, P.; Jiang, X. *SEUS: A Wearable Multi-Channel Acoustic Headset Platform to Improve Pedestrian Safety: Demo Abstract*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 330–331. [CrossRef]
332. Xia, S.; de Godoy, D.; Islam, B.; Islam, M.T.; Nirjon, S.; Kinget, P.R.; Jiang, X. A Smartphone-Based System for Improving Pedestrian Safety. In Proceedings of the 2018 IEEE Vehicular Networking Conference (VNC), Taipei, Taiwan, 5–7 December 2018; pp. 1–2. [CrossRef]
333. Lane, N.D.; Georgiev, P. Can Deep Learning Revolutionize Mobile Sensing? In Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile'15), Santa Fe, NM, USA, 12–13 February 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 117–122. [CrossRef]