# DEEP LEARNING IN HUMAN ACTIVITY RECOGNITION WITH WEARABLE SENSORS: A REVIEW ON ADVANCES *

**Shibo Zhang** *
Northwestern University
shibozhang2015@u.northwestern.edu

**Yaxuan Li** *
McGill University
yaxuanli123@gmail.com

**Shen Zhang**
Georgia Institute of Technology
shenzhang@gatech.edu

**Farzad Shahabi**
Northwestern University
farzadshahabi2024@u.northwestern.edu

**Stephen Xia**
Columbia University
sx2194@columbia.edu

**Yu Deng**
Northwestern University
yudeng2015@u.northwestern.edu

**Nabil Alshurafa**
Northwestern University
nabil@northwestern.edu

## 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. [203].

In addition to being fashion accessories, wearables provide unprecedented opportunities for monitoring human physiologic signals and facilitating natural and seamless interaction between humans and machines. Wearables integrate low-power sensors that allow them to sense movement and other physiologic signals such as heart rate, temperature, blood pressure, and electrodermal activity. The rapid proliferation of wearable technologies and advancements in sensing analytics have spurred on the growth of human activity recognition (HAR). 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® [159] 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.

*The first two authors contributed equally to this work.

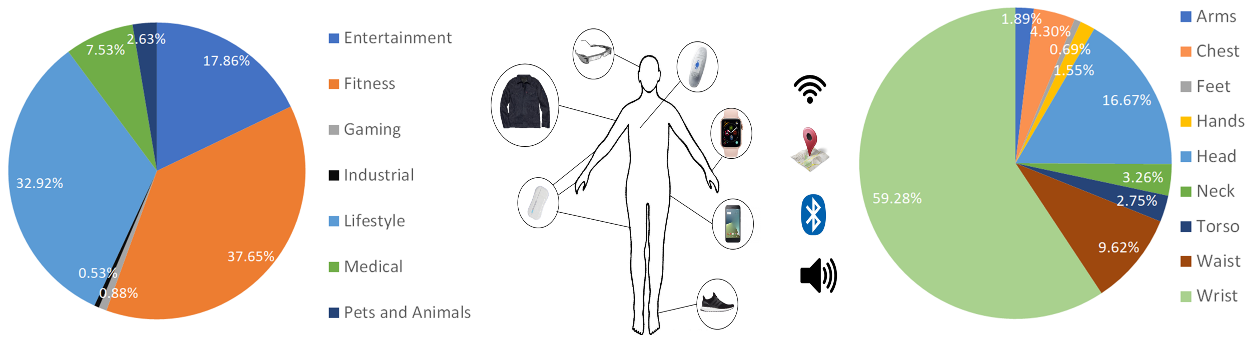


Figure 1: Wearable devices and their application. (a) Distribution of wearable applications [1]. (b) Typical wearable devices. (c) Distribution of wearable devices placed on common body areas [1].

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 [53, 169, 238]. Due to this expressive power, DL has seen an exponential 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 comprehensive survey on deep learning in HAR with wearable sensors and elaborate on ongoing challenges, obstacles, and future directions in this field.

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 [74, 188].

This article is organized as follows: Section 2.1 introduces common applications for HAR. Section 2.2 summarizes the types of sensors commonly used in HAR. Section 2.3 summarizes major datasets that are commonly used to build HAR applications. Section 3 introduces the major works in DL that contribute to HAR. Section 4 discusses major challenges, trends, and opportunities for future work. We provide concluding remarks in Section 5.

## 2 Human Activity Recognition Overview

### 2.1 Applications

In this section, we illustrate the major areas and applications of wearable devices in HAR. Fig. 1a, taken from the wearable technology database [1], breaks down the distribution of application types of 582 commercial wearables registered since 2015 [1]. 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.

#### 2.1.1 Wearables in Fitness and Lifestyle

Physical activity involves activities such as sitting, walking, laying down, going up or down stairs, jogging, and running [137]. Regular physical activity is increasingly being linked to reduction in risk for many chronic diseases, such as obesity, diabetes, and cardiovascular disease, and has been shown to improve mental health [31]. The data recorded by wearable devices during these activities include plenty of information, such as duration and intensity of activity, which further reveal an individual's daily habits and health conditions [26]. For example, dedicated products such as Fitbit [58] 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 [239]. Moreover, there has been evidence of the association between modes of transport (motor vehicle, walking, cycling, and public transport) and obesity-related outcomes [32]. 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 [30]. Therefore, the use of wearables in fitness and lifestyle has the potential to significantly advance one of the most prolific aspects of HAR applications [45, 93, 95, 113, 118, 162, 208, 209, 222].

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 habit has been an important factor in designing interventions that prevent chronic diseases such as obesity, diabetes, and cardiovascular diseases.

#### 2.1.2 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 [60, 82, 105, 135, 195, 226]. Other works have introduced methods for monitoring stroke in infants using wearable accelerometers [68] and methods for assessing depressive symptoms utilizing wrist-worn sensors [71]. 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 [133, 147, 149, 167].

#### 2.1.3 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 [174, 228, 235] 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 filed 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 [133, 149].

### 2.2 Wearable Sensors

Wearable sensors are the foundation of HAR systems. As shown in Fig. 1b, there are a large number of off-the-shelf smart devices or prototypes under development today, including smartphones, smartwatches, smart glasses, smart rings [146], smart gloves [187], smart armbands [106], smart necklaces [98], smart shoes [72], and E-tattoos [173]. These wearable devices cover the human body from head to toe with a general distribution of devices shown in Fig. 1c, as reported by [1]. 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 light weight to reduce 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.

#### 2.2.1 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 magnetic fields of the earth. 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 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 [52], Chung *et al.* have tested a range of sampling rates and given 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 [111, 112] 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.

#### 2.2.2 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 electrical activity of the heart through electrodes attached to the body. The standard 12-lead ECG attaches 10 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 the use of 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 on the signal to allow us 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.

#### 2.2.3 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 [59]. There are a number of wearable applications that detect and assess daily activities using sEMG [133, 196]. In [193], 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.

#### 2.2.4 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 [214]. Detecting these knee motions is helpful for the development of power-assisted wearables for powered lower limb protheses. The authors create a convolutional neural network and support vector machine (CNN-SVM) architecture comprising a 7-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 [132] 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) electroencephalography (EEG) for measuring brain activity, respiration sensors for breathing monitoring, ultraviolet (UV) sensors for sun exposure assessment, GPS for location sensing, microphones for audio recording, and wearable cameras for image or video recording.

It is also important to note that the wearable camera market has drastically grown with cameras such as GoPro becoming mainstream [9, 10, 81, 164] 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 [204, 218], 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.

### 2.3 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 [14, 15, 23, 24, 33, 41, 42, 99, 108, 134, 158, 175, 176, 184, 197, 200, 213, 231, 234], there are also a few datasets dedicated to hand gestures [35, 109], breathing patterns [40], and car assembly line activities [220], as well as those that monitor gait for patients with PD [16].

Table 1: Major Datasets for HAR

| Dataset | Application | Sensor | # Classes | Spl. Rate | Citations/yr |
|---|---|---|---|---|---|
| WISDM [108] | Locomotion | 3D Acc. | 6 | 20 Hz | 217 |
| ActRecTut [35] | Hand gestures | 9D IMU | 12 | 32 Hz | 153 |
| UCR(UEA)-TSC [17,44] | 9 datasets (e.g., uWave [123]) | Vary | Vary | Vary | 107* |
| UCI-HAD [15] | Locomotion | Smartphone 9D IMU | 6 | 50 Hz | 78 |
| Ubicomp 08 [200] | Home activities | Proximity sensors | 8 | N/A | 69 |
| SHO [176] | Locomotion | Smartphone 9D IMU | 7 | 50 Hz | 52 |
| UTD-MHAD1/2 [42] | Locomotion & activities | 3D Acc. & 3D Gyro. | 27 | 50 Hz | 39 |
| HHAR [184] | Locomotion | 3D Acc. | 6 | 50–200 Hz | 37 |
| Daily & Sports Activities [14] | Locomotion | 9D IMU | 19 | 25 Hz | 37 |
| MHEALTH [23,24] | Locomotion & gesture | 9D IMU & ECG | 12 | 50 Hz | 33 |
| Opportunity [41] | Locomotion & gesture | 9D IMU | 16 | 50 Hz | 32 |
| PAMAP2 [158] | Locomotion & activities | 9D IMU & HR monitor | 18 | 100 Hz | 32 |
| Daphnet [16] | Freezing of gait | 3D Acc. | 2 | 64 Hz | 30 |
| SHL [73] | Locomotion & transportation | 9D IMU | 8 | 100 Hz | 23 |
| SARD [175] | Locomotion | 9D IMU & GPS | 6 | 50 Hz | 22 |
| Skoda Checkpoint [220] | Assembly-line activities | 3D Acc. | 11 | 98 Hz | 21 |
| UniMiB SHAR [134] | Locomotion & gesture | 9D IMU | 12 | N/A | 20 |
| USC-HAD [231] | Locomotion | 3D ACC. & 3D Gyro. | 12 | 100 Hz | 20 |
| ExtraSensory [197] | Locomotion & activities | 9D IMU & GPS | 10 | 25–40 Hz | 13 |
| HASC [99] | Locomotion | Smartphone 9D IMU | 6 | 100 Hz | 11 |
| Actitracker [213] | Locomotion | 9D IMU & GPS | 5 | N/A | 6 |
| FIC [109] | Feeding gestures | 3D Acc. | 6 | 20 Hz | 5 |
| WHARF [33] | Locomotion | Smartphone 9D IMU | 16 | 50 Hz | 4 |

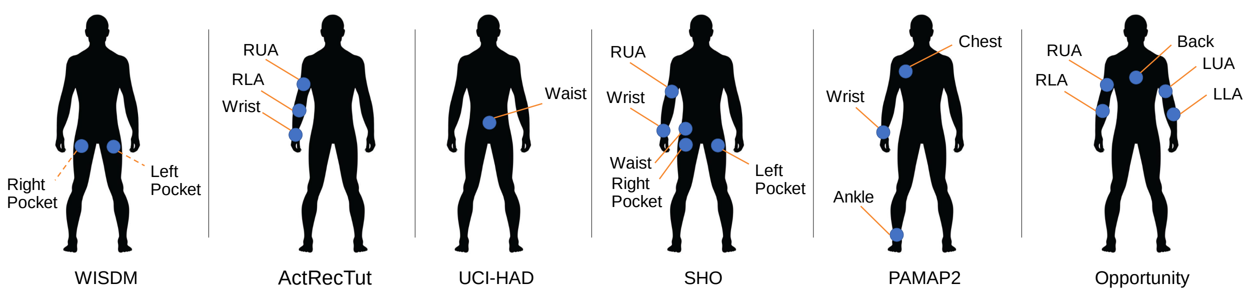


Figure 2: Placement of inertia sensors in different datasets: (a) WISDOM; (b) ActRecTut; (c) UCI-HAD; (d) SHO; (e) PAMAP2; and (f) Opportunity.

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 [55]. The UCR-TSC archive was first released and included 16 datasets, which grew to 85 datasets by 2015 and to 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 9 categories of HAR: *BasicMotions*, *Cricket*, *Epilepsy*, *ERing*, *Handwriting*, *Libras*, *NATOPS*, *RacketSports*, and *UWaveGestureLibrary* [17]. One of the most commonly used datasets is the OPPORTUNITY dataset [41]. This dataset contains data collected from 12 subjects using 15 wireless and wired networked sensor systems, with 72 sensors and 10 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 [85, 104, 117, 138, 177, 186, 191, 206, 219, 224]. Another widely-used dataset is PAMAP2 [158], 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 [220] and WISDM [108] are also commonly used to train and evaluate HAR algorithms.

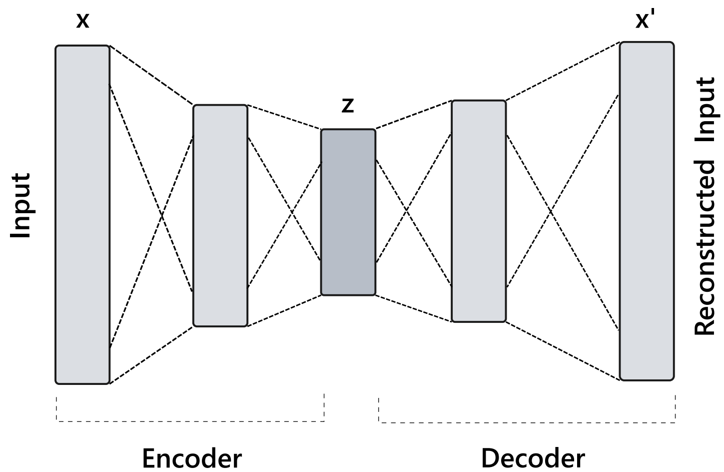


Figure 3: Illustration of an autoencoder network [166].

## 3 Deep Learning Approaches

In recent years, DL approaches have been shown to outperform 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. 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.

### 3.1 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) [22]. 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 [70, 94, 130, 163, 201]. 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 3, the encoder and decoder module is usually several dense layers (i.e., fully connected layers) of the form

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

where $\theta = \{W_e, b_e\}$, $\theta' = \{W_d, b_d\}$ are the learnable parameters of the encoder and decoder. $\sigma$ 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 [12, 18, 66, 120, 128–130, 136, 153] and functions as a powerful tool for denoising and information retrieval.

As such, autoencoders are most commonly used for feature extraction and dimensionality reduction [13, 20, 38, 49, 50, 56, 70, 94, 120, 121, 139, 163, 201, 205, 210, 212]. Autoencoders are generally used as is, 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 [120]. Experiments are performed on a public HAR dataset [15] 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 [97] studied 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 3-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 [6], detecting unseen data [102], and recognizing null classes [5].

Furthermore, autoencoders are commonly used to sanitize and denoise raw sensor data [66, 136, 198], a known problem with wearable signals that impacts our ability to learn patterns in the data. Mohammed and Tashev in [136] 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 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 [15] [66]. Then, LightGBM (LBG) is used to classify activities using the denoised signals.

Autoencoders are also commonly used to detect abnormal muscle movements, such as those resulting from Parkinson's Disease and Autism Spectrum Disorder (ASD). Rad *et al.* in [153] utilizes an autoencoder to denoise and extract optimized features of different movements and use a one-class SVM to detect movement anomalies. To reduce overfitting the autoencoder, the authors inject artificial noise, to simulate different types of perturbations, into the training data. Sigcha *et al.* in [178] uses a denoising autoencoder to detect freezing of gait (FOG) in Parkinson's disease patients. The autoencoder is only trained using data labelled as 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 [130]. 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 [41], Skoda [220], and Hand-Gesture [35] 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 non-sensitive 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. It attempts to recognize an activity without disclosing a participant's gender due to privacy concerns. The Neutralizer is an optimizer that helps the Guardian converge to a near-optimal transformation function. Both the publicly available MobiAct [202] 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%.

### 3.2 Deep Belief Network (DBN)

A DBN, as illustrated in Fig. 4, 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), which is 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.

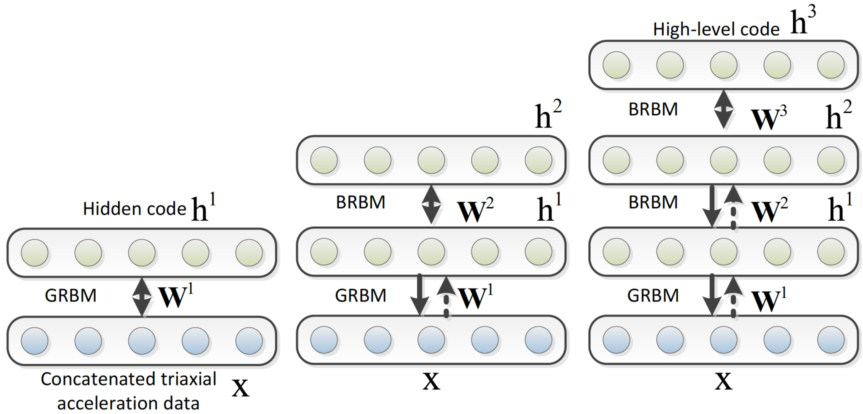


Figure 4: The greedy layer-wise training of DBNs. The first level is trained on triaxial acceleration data. Then, more RBMs are repeatedly stacked to form a deep activity recognition model [3].

The observation that DBNs can be trained greedily led to one of the first effective deep learning algorithms [27]. There are many attractive implementations and uses of DBNs in real-life applications such as drug discovery [77], natural language understanding [115], fault diagnosis [227], etc. There are also many attempts to perform HAR with DBNs. In early exploratory work back in 2011 [229], a 5-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 [3, 154, 230]. In [230], two 5-layer DBNs with different structures are applied to the Opportunity dataset [41], USCHAD dataset [231], and DSA dataset [14], and the results demonstrate improved accuracy for HAR over traditional ML methods for all the three datasets. Specifically, the accuracy for the Opportunity, USCHAD, 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.* [3] tested the activity recognition performance of DBNs using different parameter settings. Instead of using the raw acceleration data similar to [229], 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 [108], Daphnet [16], and Skoda [220] benchmark datasets, respectively. In [154], 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 [67, 83]. In [83], 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 2-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 [86] for details). In [67], a unsupervised 5-layer DBM-DNN is applied for the automatic detection of eating episodes via commercial bluetooth 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.

### 3.3 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

Table 2: Summary of studies that use CNN in HAR and their configurations. Key: C - convolutional layer, P - max-pooling layer, FC - fully connected layer, S - softmax, S1 - accelerometer, S2 - gyroscope, S3 - magnetometer

| Study | Architecture | Kernel Conv. | Application | # Classes | Sensors | Dataset |
|---|---|---|---|---|---|---|
| [114] | C-P-FC-S | 1×3, 1×4, 1×5 | locomotion activities | 3 | S1 | Self |
| [224] | C-P-FC-FC-S | 1×20 | daily activities, locomotion activities | - | - | Skoda, Opportunity, Actitracker |
| [80] | C-P-C-P-FC-FC | - | daily activities, locomotion activities | 12 | S1, S2, S3 ECG | mHealth |
| [96] | 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 |
| [46] | C-P-C-P-C-P-S | 12×2 | locomotion activities | 8 | S1 | Self |
| [219] | C-P-C-P-U-FC-S, U: unification layer | 1×3, 1×5 | daily activities, hand gesture | 18 (Opp) 12 (hand) | S1, S2 (1 for each) | Opportunity Hand Gesture |
| [193] | C-C-P-C-C-P-FC | 1×8 | hand motion classification | 10 | EMG | Rami EMG Dataset |
| [138] | C-C-P-C-C-P-FC-FC (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 |
| [131] | C-P-C-P-C-P-FC-FC-FC | 1×4, 1×10, 1×15 | locomotion activities | 6 | S1,S2,S3 | Self |

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 [168], 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 reduces feature dimensionality before it is fed into the output layer. A softmax classifier is usually selected as the final output layer. But as an exception, some studies explored the use of traditional classifiers as the output layer in a CNN [186, 214].

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 [114]. Researchers have tried encoding time-series data into 2D images as input into the CNN. In [131], 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 1-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 [7]. The choice of 1D-CNN kernel size normally falls in the range of 1×3 to 1×5 (with exceptions in [193, 214, 224] 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 5 to 9 layers [46, 62, 138, 161, 193, 214, 219], usually including 2 to 3 convolutional layers, 2 to 3 max pooling layers, followed by 1 to 2 fully connected layers before feeding the feature representation into the output layer (softmax layer in most cases). [157] used a shallow 3-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. [224] and [114] also used a small number of layers (4 layers). The choice of 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.

The number of sensors used in a HAR study can vary from a single one to as many as 23 [41]. In [46], a single accelerometer is used to collect data from 3 locations on the body: cloth pocket, trouser pocket and waist. The authors collect data on 100 subjects including 8 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, [76] proposed a multi-branch CNN architecture. A multi-branch design adopts a parallel structure which 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 7 layers, then the outputs of each branch are concatenated and fed into a fully connected and a softmax output layer.

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, [131] adapt 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.

### 3.4 Recurrent Neural Network (RNN)

Initially, the idea of using temporal information was proposed in 1991 [141] to recognize a finger alphabet consisting of 42 symbols and in 1995 [199] 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 [37, 43, 90, 124, 189, 190, 194].

Unlike feed-forward neural networks, a RNN processes the input data in a recurrent behavior. Equivalent to a directed graph, RNN exhibits dynamic behaviors and possesses the capability of modeling temporal and sequential relationships due to a hidden layer with recurrent connections. A typical structure for an RNN is shown in Figure 5 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:

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

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) [237], Continuous Time RNN (CTRNN) [19], Personalized RNN (PerRNN) [211], Colliding Bodies Optimization RNN (CBO-RNN) [103]. Unlike previous models with 1-dimension time-series input, [126] builds a CNN+RNN model with stacked multisensor data in each channel for fusion before feeding into CNN layer. [101] 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 [28]. In an effort to address these challenges, long short-term memory (LSTM)-based RNNs [88], and gated recurrent units (GRU) [51] 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 [54, 89, 92, 144, 240]. The LSTM has been shown capable of memorizing and modeling 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 the field of human activity recognition, speech recognition, handwriting recognition, natural language processing, video analysis, etc. As illustrated in Figure 5 [140], 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 [140].

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

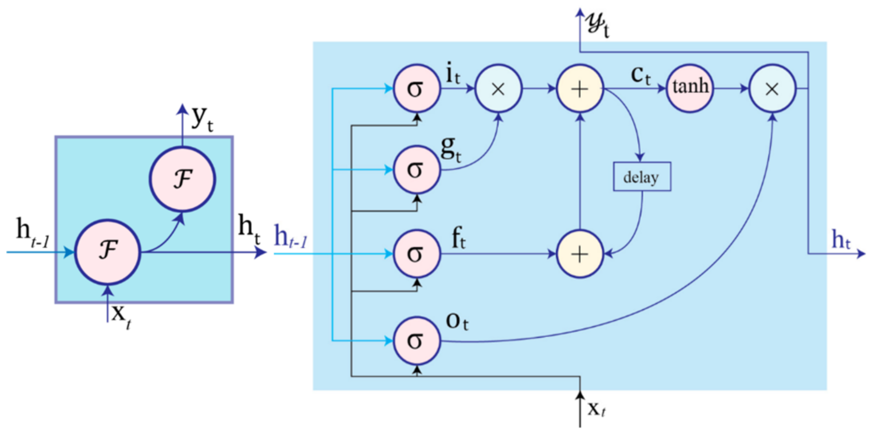


Figure 5: Schematic diagram of an RNN node and LSTM cell [140]. Left: RNN node where $h_{t-1}$ is the previous hidden state, $x_t$ is the current input sample data, $h_t$ is the current hidden state, $y_t$ is the current output, and $\mathscr{F}$ is the activation function. Right: LSTM cell with internal recurrence $c_t$ and outer recurrence $h_t$.

As shown in Figure 6, 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 [140, 144, 160]. Researchers in [84] 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 [84]. Bidirectional LSTMs, having both past and future recurrent connections, were used in [116, 192] to classify activities.

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, [89] proposes a harmonic loss function and [221] combines LSTM with batch normalization to achieve 92% accuracy with raw accelerometer and gyroscope data. [145] 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 [25] 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 [151]. They achieved 98.1% accuracy on SHL transportation modes classification dataset with CNN-extracted and hand-crafted features as input. Similarly, other researchers [4, 57, 64, 65, 142, 170, 215, 216] have also developed the CNN-LSTM model in varios application scenarios by taking advantage of the feature extraction ability of CNN and the time-series data reasoning ability of LSTM.

Raw IMU and EMG time series data are commonly used as inputs to RNNs [34, 47, 48, 79, 87, 101]. A number of major datasets used to train and evaluate RNN models have been created, including the Sussex-Huawei Locomotion-Transportation (SHL) [237, 240], PAMAP2 [126, 223] and Opporunity [160]. In addition to raw time series data [144], Besides raw time series data, custom features are also commonly used as inputs to RNNs. [54] showed that training a 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 build direct mappings

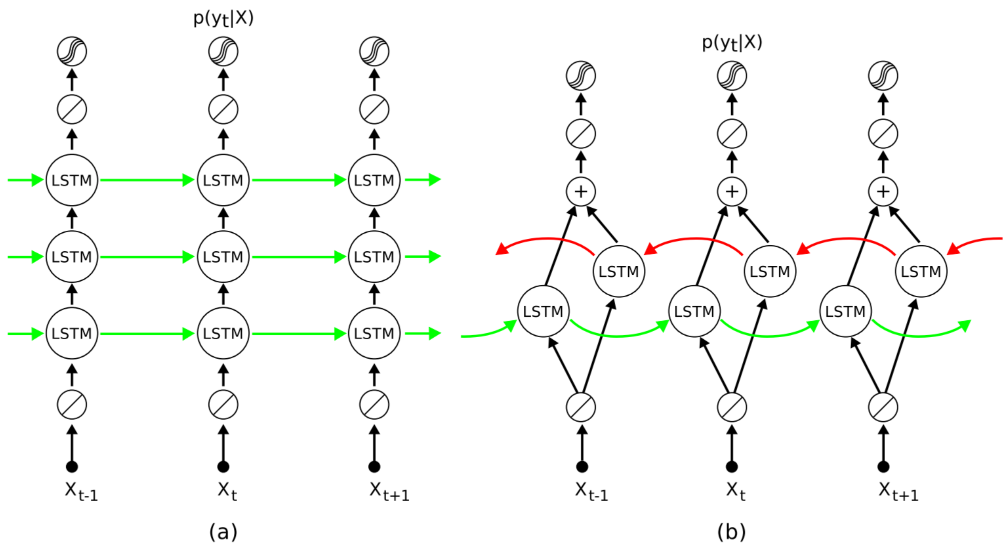


Figure 6: The structure of LSTM and bi-directional LSTM model [84]. (a). LSTM network hidden layers containing LSTM cells and a final softmax layer at the top. (b) bi-directional LSTM network with two parallel tracks in both future (green) and past (red) directions.

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 centered 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. [92] utilizes attention in combination with a 1D CNN Gated Recurrent Units (GRU), achieving HAR performances of 96.5% ± 1.0%, 93.1% ± 2.2%, and 89.3% ± 1.3% on Heterogeneous [185], Skoda [220], and PAMAP2 [158] datasets, respectively. [223] 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 are 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 [217], 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).

### 3.5 Deep Reinforcement Learning (DRL)

AE, DBN, CNNs 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, the three key elements - action, state and reward collectively form the paradigm of RL. The structure of RL is shown in Fig. 7.

In the domain of HAR, [171] uses DRL to predict arm movements with 98.33% accuracy. [236] 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 [29], Bhat et al. propose an HAR online learning framework which takes advantage of reinforcement learning utilizing policy gradient algorithm for faster convergence achieving 97.7% in recognizing 6 activities.

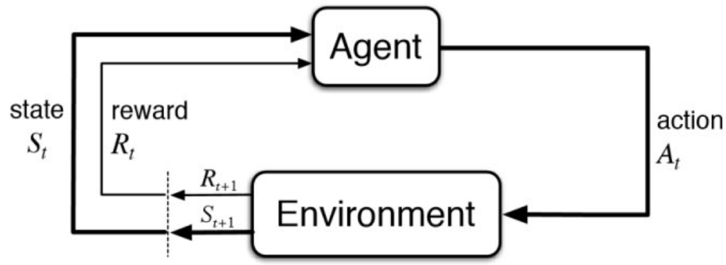


Figure 7: A typical structure of a reinforcement learning network [107].

### 3.6 Deep Generative Adversarial Network (DGAN)

Originally proposed to generate credible fake images that resembles 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 [75]. It comprises two networks, the generator ($\mathbf{G}$) and the discriminator ($\mathbf{D}$), competing against each other in a zero-sum game framework as shown in Fig. 8. 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 $\hat{\mathbf{x}}$ to challenge the discriminator to differentiate between original samples $\mathbf{x}$ and fake samples $\hat{\mathbf{x}}$. 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 game zero-sum 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 [61].

GAN model has shown remarkable performance in generating synthetic data with high quality and rich details. In the field of HAR, GAN has been applied as a semi-supervised learning approach to deal with unlabeled or partially labeled 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 [127, 143, 181]. Afterward, GAN has shown the ability to generate balanced and realistic synthetic sensor data. [207] utilizes GANs with a customized network to generate synthetic data from the public HAR dataset HASC2010corpus [100]. Similarly, [8] assesses 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 is built with CNN layers, and the results demonstrate synthetic data with high quality and diversity with two public datasets. Moreover, by oversampling and adding synthetic sensor data into the training, researchers augment and alleviate the originally imbalanced training set to achieve better performance. [39, 119] generate verisimilar data of different activities and [172] uses the Boulic kinematic model which aims to capture the three dimensional positioning trend to synthesize personified walking data. Moreover, [180] is an attempt that utilizes GAN to perform cross-subject transfer learning for HAR since collecting data for each new user is infeasible. 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.

## 4 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 three questions, which overlap with the three major components of machine learning.

- Q1 : What are the challenges in data acquisition? How do we resolve them?
- Q2 : What are the challenges in label acquisition? What are the current methods for them?
- Q3 : What are the challenges in modeling? What are potential solutions?

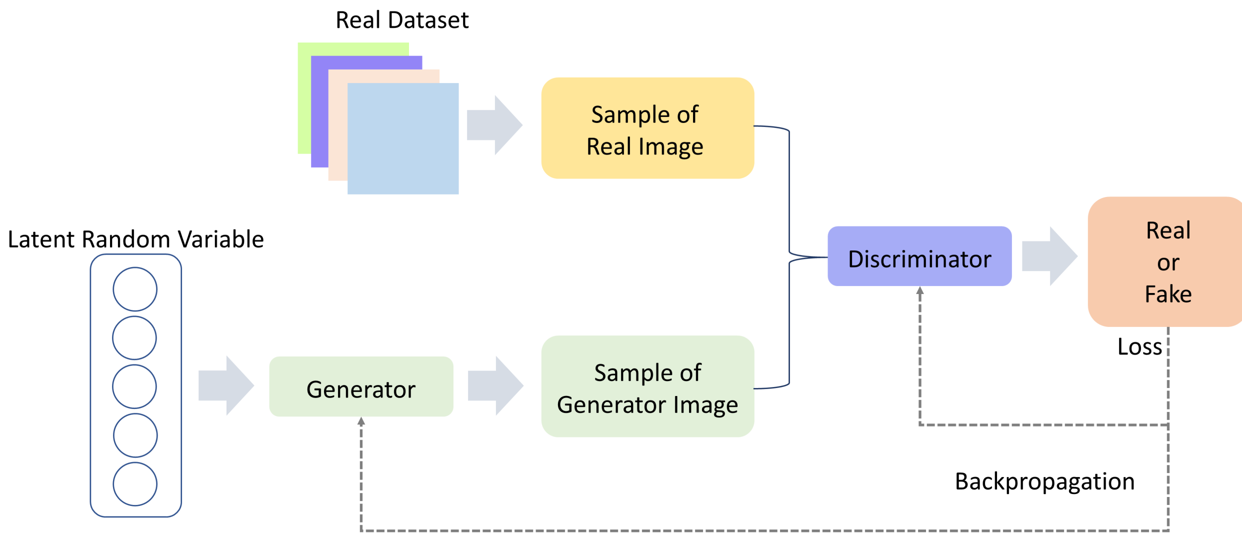


Figure 8: The structure of generative adversarial network.

### 4.1 Challenges in Data Acquisition and Sample Size

Data is the cornerstone of artificial intelligence. Models only perform as well as the quality of the data used during training. If the expectation is to ensure that the data is generalizable, then careful attention should be made in data collection to ensure the participants are representative of the population of interest. Moreover, showing sufficient sample size is an important problem. In HAR there currently is no well-defined method for determining sample size of training data, however, showing convergence of the error rate as a function of the number of samples in the training data is one approach shown by Yang *et al.* [69]. Acquiring a massive amount of high-quality data at a low cost is a critical problem in every domain. There have been extensive efforts to collect and create publicly available datasets in some domains, such as computer vision and natural language processing. As such, researchers in these domains have access to ample amounts of data. However, in HAR, collecting raw data is labor-intensive considering a large number of different sensors and types of wearables that could be used. Therefore, proposing and developing innovative approaches to augmenting data with synthetic data all while ensuring high-quality data collection is imperative for the growth of HAR research. **Opportunity:** To address this challenge, research in methods to simplify or improve HAR sensors and their robust performance in real-world scenarios are needed to enable modeling of humans in the real-world.

#### 4.1.1 The Need for More Data

Raw data collection requires a considerable amount of time and effort for projects 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 data during training when there is a shortage of collected data. Data augmentation applies synthetic noise, typically observed in real situations, to real data to obtain new training samples. In general, using the dataset augmented with synthetic training samples yield higher classification accuracy when compared to using the same dataset without augmentation. 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 [72]. Many DL-based methods also augment their datasets to improve performance [91, 156, 182, 195]. **Opportunity:** To address this challenge, research is needed in quantifying the benefits of different data augmentation methods to assess their benefits and limitations as a function of different data sets.

#### 4.1.2 Data Quality and Missing Data

The quality of models is highly dependent on the quality of the data used for training. Many real-world collection scenarios introduce different sources of noise that may significantly degrade data quality, such as electromagnetic interference or uncertainty in task scheduling for devices that perform sampling [155].

In addition to developing robust noise-resistant systems for data collection, 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 [36]. Luo *et al.* utilized a GAN to infer missing time series data [125]. Saeed *et al.* proposed an adversarial autoencoder (AAE) framework to perform data imputation [165]. **Opportunity:** To address this challenge, more research into automated methods for evaluating and quantifying the quality is needed in order to better identify, remove, and/or correct for poor quality data.

### 4.2 Challenges in Label Acquisition

Labeled data is crucial for deep supervised learning. Image and audio data is generally easy to label by visually confirming or hearing the images or audio streams, respectively. However, labeling and distinguishing different types of human activities simply by looking at different time series from HAR sensors is difficult as unique patterns in raw data can be difficult to identify, especially as the number of potential confounding activities increases. Therefore, ground truth acquisition for HAR sensors generally requires an extra wearable or sensing source to provide video or audio data combined with synchronization methods to align timestamps from the ground truth video/audio with the HAR sensor data. This makes ground truth acquisition for HAR much more labor-intensive. Moreover, accurate time synchronization between wearable devices and video cameras is challenging and an emerging research problem in ubiquitous computing because different devices are equipped with independent (and often drifting) clocks. A number of attempts have been made to address this issue, such as [63] and the Window Induced Shift Estimation method (SyncWISE) [233]. **Opportunity:** Greater research methods are needed that integrate expert knowledge for automated labeling of human activities.

#### 4.2.1 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.

1. Semi-supervised Learning

   Semi-supervised learning utilizes both labeled data and unlabeled data. By leveraging the abundance of unlabeled data, it is able to learn a more generalizable feature representation. Zeng *et al.* presented two semi-supervised CNN methods that utilizes unlabeled data during training: the convolutional encoder-decoder and the convolutional ladder network [225]. They showed that the convolutional ladder network had a better classification performance, achieving 18% F1-score improvement compared to the supervised CNN approach on 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 [21].

2. Active Learning

   Active learning is an advanced ground-truthing technique that, based on an objective function, selectively chooses unlabeled data for a human-in-the-loop annotator to label. In active learning, the model predicts the label and generates a confidence measurement. Samples with low prediction confidence are sent to human annotators to provide labels. In recent years, researchers have tried to combine DL approaches with an active learning framework, in order to gain the benefit of 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) [78]. **Opportunity:** Though active learning has demonstrated that fewer labels are needed to build an effective model, it is unknown if this method of working with an oracle to select the best data samples to segment truly saves annotator time and cost. A time-cost analysis is needed with varying datasets to truly determine the benefits of active learning. Moreover, given the many existing labeled datasets, another area of opportunity is developing methods that leverage characteristics of labeled datasets to generate labels for unlabeled datasets (i.e., transfer learning).

#### 4.2.2 Privacy Protection

With the growing number and inference potential of novel sensors, privacy is increasingly becoming a concern among users. In general, the more inference potential a sensor has, the less willing a person is to agree to its collection of data, not at-least without some measure of control. To address this issue, multiple works have introduced methods for preserving user privacy 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.

#### 4.2.3 Issues of In-the-field Dataset

Traditionally, HAR research has been conducted primarily in-lab. Recently, HAR research has been moving towards in-field experimentation. Unlike in-lab settings, where the ground truth of activities can be collected with surveillance cameras and annotated by human annotators, in-field experiments may have people moving around in daily life. A static camera deployment would not be sufficient to capture all the places a person may go. Alharbi *et al*. used ground truth wearable cameras placed at the wrist, chest, and shoulders to record subject movements and surroundings as subjects moved around outside of a lab setting [11]. **Opportunity:** More research in leveraging human-in-the-loop to provide in-field labeling is required to generate more robust datasets for in situ activities. 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.

### 4.3 Challenges in Modeling

#### 4.3.1 Data Segmentation

As discussed in [150], many methods segment time series using traditional static sliding window methods. A static time window may either be too large, capturing more than is necessary to detect certain activities, or too small and not capturing enough of the series to detect long movements and activities. 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 [232]. **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.

#### 4.3.2 Semantically Complex Activity Recognition

Current HAR methods achieve high performance for simple activities such as running. However, complex activities such as eating, that could be composed of many different movements, is still challenging. 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 [110]. Liu *et al*. proposes a hierarchical architecture that constructs high-level human activities from low-level activities [122]. 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 activities [148]. **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 rather than specifically identified or labeled by designers, have yet to be fully explored; such a paradigm has the potential to achieve higher accuracy than methods where humans identify the higher-level attributes to learn and classify, much like the relationship between deep learning (where features are automatically learned) and classical machine learning (where features are engineered).

#### 4.3.3 Model Generalizability

A model has high generalizability when the model performs well on data that it has never seen before. Overfitting occurs when it performs well on training data, but poorly on new data. K-fold cross validation or leave-one-participant-out cross validation is generally used to improve the generalizability of the model. In wearable based HAR, many factors will affect the generalizability of the model; we group them into two categories: *User Variability* and *Sensing Heterogeneity*. During recent years, many efforts have been put into improving the generalizability of models in HAR [2, 152, 183]. Most research on generalizability in HAR has been focused on creating models that can generalize to a larger population. Creating highly generalizable models generally requires a large amount of data and high model complexity. We would like to point out that another direction is to adaptively adjust model parameters to new users as they use the system. Rather than having a model that has many parameters and is intensive to train, we have a smaller model that adaptively adjusts to the current user using this model. Siirtola and Röning propose an online incremental learning approach that continuously adapts the model with the user's individual data as it comes in [179]. **Opportunity:** However, combining incremental learning with DL in wearable HAR systems remains an open problem.

## 5 Conclusion

Human activity recognition in wearables has provided us with many conveniences and avenues to monitor and improve our quality of life. 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. This paper systematically categorizes and summarizes existing work that apply DL approaches to wearable sensors based HAR and provides comprehensive analysis of the current advancements, major barriers, developing trends, cutting-edge frontiers, and potential future directions for DL-based HAR.

## Acknowledgments

Special thanks to Haik Kalamtarian and Krystina Neuman for their valuable feedback.

## References


[1] Wearable Technology Database. [Online]. Available: `https://data.world/crowdflower/wearable-technology-database`,. Accessed: July 2021.

[2] Mohamed Abdel-Basset, Hossam Hawash, Victor Chang, Ripon K. Chakrabortty, and Michael Ryan. Deep learning for heterogeneous human activity recognition in complex iot applications. *IEEE Internet of Things Journal*, pages 1–1, 2020.

[3] Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan. Deep Activity Recognition Models with Triaxial Accelerometers. *arXiv e-prints*, page arXiv:1511.04664, Nov 2015.

[4] Waqar Ahmad, Bibi Misbah Kazmi, and Hazrat Ali. Human activity recognition using multi-head cnn followed by lstm. In *2019 15th international conference on emerging technologies (ICET)*, pages 1–6. IEEE, 2019.

[5] Ali Akbari and Roozbeh Jafari. An autoencoder-based approach for recognizing null class in activities of daily living in-the-wild via wearable motion sensors. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3392–3396. IEEE, 2019.

[6] Ali Akbari and Roozbeh Jafari. 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*, pages 85–96, 2019.

[7] Ali Akbari and Roozbeh Jafari. Transition-aware detection of modes of locomotion and transportation through hierarchical segmentation. *IEEE Sensors Journal*, 21(3):3301–3313, 2020.

[8] Fayez Alharbi, Lahcen Ouarbya, and Jamie A Ward. Synthetic sensor data for human activity recognition. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2020.

[9] Rawan Alharbi, Tammy Stump, Nilofar Vafaie, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. I can't be myself: effects of wearable cameras on the capture of authentic behavior in the wild. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2(3):1–40, 2018.

[10] Rawan Alharbi, Mariam Tolba, Lucia C Petito, Josiah Hester, and Nabil Alshurafa. To mask or not to mask? balancing privacy with visual confirmation utility in activity-oriented wearable cameras. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 3(3):1–29, 2019.

[11] Rawan Alharbi, Nilofar Vafaie, Kitty Liu, Kevin Moran, Gwendolyn Ledford, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. Investigating barriers and facilitators to wearable adherence in fine-grained eating detection. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 407–412. IEEE, 2017.

[12] Bandar Almaslukh, Jalal AlMuhtadi, and Abdelmonim Artoli. An effective deep autoencoder approach for online smartphone-based human activity recognition. *Int. J. Comput. Sci. Netw. Secur*, 17:160, 2017.

[13] Bandar Almaslukh, Jalal AlMuhtadi, and Abdelmonim Artoli. An effective deep autoencoder approach for online smartphone-based human activity recognition. *Int. J. Comput. Sci. Netw. Secur*, 17(4):160–165, 2017.

[14] Kerem Altun, Billur Barshan, and Orkun Tunçel. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recogn.*, 43(10):3605–3620, Oct 2010.

[15] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *Esann*, 2013.

[16] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster. Wearable assistant for parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):436–446, March 2010.

[17] Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018.

[18] Lu Bai, Chris Yeung, Christos Efstratiou, and Moyra Chikomo. Motion2vector: unsupervised learning in human activity recognition using wrist-sensing data. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 537–542, 2019.

[19] Gonzalo Bailador, Daniel Roggen, Gerhard Tröster, and Gracián Triviño. Real time gesture recognition using continuous time recurrent neural networks. In *BodyNets*, page 15. Citeseer, 2007.

[20] Dmitrijs Balabka. Semi-supervised learning for human activity recognition using adversarial autoencoders. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 685–688, 2019.

[21] Dmitrijs Balabka. Semi-supervised learning for human activity recognition using adversarial autoencoders. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, page 685–688, New York, NY, USA, 2019. Association for Computing Machinery.

[22] Dana H. Ballard. Modular learning in neural networks. In *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, AAAI'87, pages 279–284. AAAI Press, 1987.

[23] Oresti Banos, Rafael Garcia, Juan A Holgado-Terriza, Miguel Damas, Hector Pomares, Ignacio Rojas, Alejandro Saez, and Claudia Villalonga. mhealthdroid: a novel framework for agile development of mobile health applications. In *International Workshop on Ambient Assisted Living*, pages 91–98. Springer, 2014.

[24] Oresti Banos, Claudia Villalonga, Rafael Garcia, Alejandro Saez, Miguel Damas, Juan A Holgado-Terriza, Sungyong Lee, Hector Pomares, and Ignacio Rojas. Design, implementation and validation of a novel open framework for agile development of mobile health applications. *Biomedical engineering online*, 14(2):S6, 2015.

[25] Onur Barut, Li Zhou, and Yan Luo. Multitask lstm model for human activity recognition and intensity estimation using wearable sensor data. *IEEE Internet of Things Journal*, 7(9):8760–8768, 2020.

[26] Adrian E Bauman, Rodrigo S Reis, James F Sallis, Jonathan C Wells, Ruth JF Loos, and Brian W Martin. Correlates of physical activity: why are some people physically active and others not? *The Lancet*, 380(9838):258 – 271, 2012.

[27] Yoshua Bengio. Learning deep architectures for ai. *Foundations and Trends® in Machine Learning*, 2(1):1–127, 2009.

[28] Yoshua Bengio, Paolo Frasconi, J urgen Schmidhuber, and Corso Elvezia. Gradient flow in recurrent nets: the difficulty of learning long-term d ependencies* sepp hochreiter fakult at f ur informatik.

[29] Ganapati Bhat, Ranadeep Deb, Vatika Vardhan Chaurasia, Holly Shill, and Umit Y Ogras. Online human activity recognition using low-power wearable devices. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–8. IEEE, 2018.

[30] Alycia Bisson and Margie E. Lachman. Behavior change with fitness technology in sedentary adults: A review of the evidence for increasing physical activity. *Frontiers in Public Health*, 4, 01 2017.

[31] Frank W Booth, Christian K Roberts, and Matthew J Laye. Lack of exercise is a major cause of chronic diseases. *Comprehensive Physiology*, 2(2):1143–1211, 2011.

[32] V Brown, M Moodie, AM Mantilla Herrera, JL Veerman, and R Carter. Active transport and obesity prevention–a transportation sector obesity impact scoping review and assessment for melbourne, australia. *Preventive medicine*, 96:49–66, 2017.

[33] B. Bruno, F. Mastrogiovanni, and A. Sgorbissa. A public domain dataset for adl recognition using wrist-placed accelerometers. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 738–743, Aug 2014.

[34] Nan Bu, Osamu Fukuda, and Toshio Tsuji. Emg-based motion discrimination using a novel recurrent neural network. *Journal of Intelligent Information Systems*, 21(2):113–126, 2003.

[35] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.*, 46(3):33:1–33:33, Jan 2014.

[36] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 6775–6785. Curran Associates, Inc., 2018.

[37] Alessandro Carfì, Carola Motolese, Barbara Bruno, and Fulvio Mastrogiovanni. Online human gesture recognition using recurrent neural networks and wearable sensors. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 188–195. IEEE, 2018.

[38] M. P. Centeno, A. v. Moorsel, and S. Castruccio. Smartphone continuous authentication using deep learning autoencoders. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, pages 147–1478, 2017.

[39] Mang Hong Chan and Mohd Halim Mohd Noor. A unified generative model using generative adversarial network for activity recognition. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10, 2020.

[40] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. Breathprint: Breathing acoustics-based user authentication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '17, pages 278–291, New York, NY, USA, 2017. ACM.

[41] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R. Millán, and Daniel Roggen. The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033 – 2042, 2013. Smart Approaches for Human Action Recognition.

[42] C. Chen, R. Jafari, and N. Kehtarnavaz. Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 168–172, Sep. 2015.

[43] Weitong Chen, Sen Wang, Xiang Zhang, Lina Yao, Lin Yue, Buyue Qian, and Xue Li. Eeg-based motion intention recognition via multi-task rnns. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 279–287. SIAM, 2018.

[44] Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, and Gustavo Batista. The ucr time series classification archive, July 2015.

[45] Yuqing Chen and Yang Xue. A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1488–1492, Oct 2015.

[46] Yuqing Chen and Yang Xue. A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE international conference on systems, man, and cybernetics*, pages 1488–1492. IEEE, 2015.

[47] Guy Cheron, Ana Maria Cebolla, Ana Bengoetxea, Françoise Leurs, and Bernard Dan. Recognition of the physiological actions of the triphasic emg pattern by a dynamic recurrent neural network. *Neuroscience letters*, 414(2):192–196, 2007.

[48] Guy Chéron, Françoise Leurs, Ana Bengoetxea, JP Draye, Maud Destrée, and Bernard Dan. A dynamic recurrent neural network for multiple muscles electromyographic mapping to elevation angles of the lower limb in human locomotion. *Journal of neuroscience methods*, 129(2):95–104, 2003.

[49] Belkacem Chikhaoui and Frank Gouineau. Towards automatic feature extraction for activity recognition from wearable sensors: a deep learning approach. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 693–702. IEEE, 2017.

[50] Edison A Chung and Marco E Benalcázar. Real-time hand gesture recognition model using deep learning techniques and emg signals. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE, 2019.

[51] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555, 2014.

[52] S. Chung, J. Lim, K. J. Noh, G. Gue Kim, and H. T. Jeong. Sensor positioning and data acquisition for activity recognition using deep learning. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 154–159, Oct 2018.

[53] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[54] Krzysztof Czuszyński, Jacek Rumiński, and Alicja Kwaśniewska. Gesture recognition with the linear optical sensor and recurrent neural networks. *IEEE Sensors Journal*, 18(13):5429–5438, 2018.

[55] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive, 2018.

[56] Fernando Henrique Cruz De Andrade, Flavio Garcia Pereira, Cassius Zanetti Resende, and Daniel Cruz Cavalieri. Improving semg-based hand gesture recognition using maximal overlap discrete wavelet transform and an autoencoder neural network. In *XXVI Brazilian Congress on Biomedical Engineering*, pages 271–279. Springer, 2019.

[57] Samundra Deep and Xi Zheng. Hybrid model featuring cnn and lstm architecture for human activity recognition on smartphone sensor data. In *2019 20th international conference on parallel and distributed computing, applications and technologies (PDCAT)*, pages 259–264. IEEE, 2019.

[58] Keith M. Diaz, David J. Krupka, Melinda J. Chang, James Peacock, Yao Ma, Jeff Goldsmith, Joseph E. Schwartz, and Karina W. Davidson. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, 185:138–140, Apr 2015.

[59] Electromyography. Electromyography — Wikipedia, the free encyclopedia, 2010. [Online; accessed 10-March-2020].

[60] B. M. Eskofier, S. I. Lee, J. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz, and P. Bonato. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 655–658, Aug 2016.

[61] Farzan Farnia and Asuman Ozdaglar. Gans may have no nash equilibria. *arXiv preprint arXiv:2002.09124*, 2020.

[62] Yanggang Feng, Wanwen Chen, and Qining Wang. A strain gauge based locomotion mode recognition method using convolutional neural network. *Advanced Robotics*, 33(5):254–263, 2019.

[63] Lex Fridman, Daniel E. Brown, William Angell, Irman Abdic, Bryan Reimer, and Hae Young Noh. Automated synchronization of driving data using vibration and steering events. *CoRR*, abs/1510.06113, 2015.

[64] Björn Friedrich, Carolin Lübbe, and Andreas Hein. Combining lstm and cnn for mode of transportation classification from smartphone sensors. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pages 305–310, 2020.

[65] Jing Gao, Peishang Gu, Qing Ren, Jinde Zhang, and Xin Song. Abnormal gait recognition algorithm based on lstm-cnn fusion network. *IEEE Access*, 7:163180–163190, 2019.

[66] Xile Gao, Haiyong Luo, Qu Wang, Fang Zhao, Langlang Ye, and Yuexia Zhang. A human activity recognition algorithm based on stacking denoising autoencoder and lightgbm. *Sensors*, 19(4), 2019.

[67] Y. Gao, N. Zhang, H. Wang, X. Ding, X. Ye, G. Chen, and Y. Cao. ihear food: Eating detection using commodity bluetooth headsets. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 163–172, June 2016.

[68] Yan Gao, Yang Long, Yu Guan, Anna Basu, Jessica Baggaley, and Thomas Ploetz. Towards reliable, automated general movement assessment for perinatal stroke screening in infants using wearable accelerometers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(1):12:1–12:22, March 2019.

[69] Yang Gao, Yincheng Jin, Jagmohan Chauhan, Seokmin Choi, Jiyang Li, and Zhanpeng Jin. Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(1), March 2021.

[70] Y. Gavrilin and A. Khan. Across-sensor feature learning for energy-efficient activity recognition on mobile devices. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2019.

[71] A. Ghandeharioun, S. Fedor, L. Sangermano, D. Ionescu, J. Alpert, C. Dale, D. Sontag, and R. Picard. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 325–332, Oct 2017.

[72] Giacomo Giorgi, Fabio Martinelli, Andrea Saracino, and Mina Sheikhalishahi. Try walking in my shoes, if you can: Accurate gait recognition through deep learning. In Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch, editors, *Computer Safety, Reliability, and Security*, pages 384–395, Cham, 2017. Springer International Publishing.

[73] H. Gjoreski, M. Ciliberto, L. Wang, F. J. Ordonez Morales, S. Mekki, S. Valentin, and D. Roggen. The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access*, 6:42592–42604, 2018.

[74] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. 2016. `http://www.deeplearningbook.org`.

[75] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[76] Rene Grzeszick, Jan Marius Lenk, Fernando Moya Rueda, Gernot A Fink, Sascha Feldhorst, and Michael ten Hompel. 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*, pages 1–6, 2017.

[77] Wanrong Gu, Guohua Wang, Ziye Zhang, Yijun Mao, Xianfen Xie, and Yichen He. High accuracy drug-target protein interaction prediction method based on dbn. In *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pages 58–62. IEEE, 2020.

[78] Gautham Krishna Gudur, Prahalathan Sundaramoorthy, and Venkatesh Umaashankar. Activeharnet: Towards on-device deep bayesian active learning for human activity recognition. In *The 3rd International Workshop on Deep Learning for Mobile Systems and Applications*, EMDL '19, page 7–12, New York, NY, USA, 2019. Association for Computing Machinery.

[79] Rohit Gupta, Inderjeet Singh Dhindsa, and Ravinder Agarwal. Continuous angular position estimation of human ankle during unconstrained locomotion. *Biomedical Signal Processing and Control*, 60:101968, 2020.

[80] Sojeong Ha and Seungjin Choi. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 381–388. IEEE, 2016.

[81] Aksasse Hamid, Aksasse Brahim, Ouanan Mohammed, et al. A survey of activity recognition in egocentric lifelogging datasets. In *2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, pages 1–8. IEEE, 2017.

[82] Nils Y. Hammerla, James M. Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plotz. Pd disease state assessment in naturalistic environments using deep learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 1742–1748. AAAI Press, 2015.

[83] Nils Y. Hammerla, James M. Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plotz. Pd disease state assessment in naturalistic environments using deep learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 1742–1748. AAAI Press, 2015.

[84] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880*, 2016.

[85] Mohammed Mehedi Hassan, Md Zia Uddin, Amr Mohamed, and Ahmad Almogren. A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, 81:307–313, 2018.

[86] Geoffrey E Hinton. A practical guide to training restricted boltzmann machines. In *Neural networks: Tricks of the trade*, pages 599–619. Springer, 2012.

[87] Masaaki Hioki and Haruhisa Kawasaki. Estimation of finger joint angles from semg using a recurrent neural network with time-delayed input vectors. In *2009 IEEE International Conference on Rehabilitation Robotics*, pages 289–294. IEEE, 2009.

[88] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[89] Yue Hu, Xiao-Qing Zhang, Li Xu, Feng Xian He, Zhao Tian, Wei She, and Wei Liu. Harmonic loss function for sensor-based human activity recognition based on lstm recurrent neural networks. *IEEE Access*, 8:135617–135627, 2020.

[90] Masaya Inoue, Sozo Inoue, and Takeshi Nishida. Deep recurrent neural network for mobile human activity recognition with high throughput. *Artificial Life and Robotics*, 23(2):173–185, 2018.

[91] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Data augmentation using synthetic data for time series classification with deep residual networks. In *International Workshop on Advanced Analytics and Learning on Temporal Data, ECML PKDD*, 2018.

[92] Mahendra Kumar Jangir and Karan Singh. Hargrurnn: Human activity recognition using inertial body sensor gated recurrent units recurrent neural network. *Journal of Discrete Mathematical Sciences and Cryptography*, 22(8):1577–1587, 2019.

[93] Jeya Vikranth Jeyakumar, Eun Sun Lee, Zhengxu Xia, Sandeep Singh Sandha, Nathan Tausik, and Mani Srivastava. 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 '18, pages 1606–1615, New York, NY, USA, 2018. ACM.

[94] Guangyu Jia, Hak-Keung Lam, Junkai Liao, and Rong Wang. Classification of electromyographic hand gesture signals using machine learning techniques. *Neurocomputing*, 2020.

[95] Wenchao Jiang and Zhaozheng Yin. Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 1307–1310, New York, NY, USA, 2015. ACM.

[96] Artur Jordao, Leonardo Antônio Borges Torres, and William Robson Schwartz. Novel approaches to human activity recognition based on accelerometer data. *Signal, Image and Video Processing*, 12(7):1387–1394, 2018.

[97] Kyungkoo Jun and Soonpil Choi. Unsupervised end-to-end deep model for newborn and infant activity recognition. *Sensors*, 20(22):6467, 2020.

[98] Haik Kalantarian, Nabil Alshurafa, Tuan Le, and Majid Sarrafzadeh. Monitoring eating habits using a piezoelectric sensor-based necklace. *Computers in Biology and Medicine*, 58:46 – 55, 2015.

[99] Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio. 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, pages 27:1–27:5, New York, NY, USA, 2011. ACM.

[100] Nobuo Kawaguchi, Ying Yang, Tianhui Yang, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio. Hasc2011corpus: Towards the common ground of human activity recognition. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, page 571–572, New York, NY, USA, 2011. Association for Computing Machinery.

[101] István Ketykó, Ferenc Kovács, and Krisztián Zsolt Varga. Domain adaptation for semg-based gesture recognition with recurrent neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2019.

[102] Md Abdullah Al Hafiz Khan and Nirmalya Roy. Untran: Recognizing unseen activities with unlabeled data using transfer learning. In *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 37–47. IEEE, 2018.

[103] Pankaj Khatiwada, Matrika Subedi, Ayan Chatterjee, and Martin Wulf Gerdes. Automated human activity recognition by colliding bodies optimization-based optimal feature selection with recurrent neural network. *arXiv preprint arXiv:2010.03324*, 2020.

[104] Eunji Kim. Interpretable and accurate convolutional neural networks for human activity recognition. *IEEE Transactions on Industrial Informatics*, 16(11):7190–7198, 2020.

[105] Han Byul Kim, Woong Woo Lee, Aryun Kim, Hong Ji Lee, Hye Young Park, Hyo Seon Jeon, Sang Kyong Kim, Beomseok. Jeon, and Kwang S. Park. Wrist sensor-based tremor severity quantification in parkinson's disease using convolutional neural network. *Computers in Biology and Medicine*, 95:140 – 146, 2018.

[106] J. Kim, M. Kim, and K. Kim. Development of a wearable hci controller through semg & imu sensor fusion. In *2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 83–87, Aug 2016.

[107] Huang Kung-Hsiang (Steeve). Introduction to Various Reinforcement Learning Algorithms. Part I (Q-Learning, SARSA, DQN, DDPG). `https://towardsdatascience.com/introduction-to-various-reinforcement-learning-algorithms-i-q-learning-sarsa-dqn-ddpg-72a5e0cb6287`, 2018. [Online; accessed 13-July-2021].

[108] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, Mar 2011.

[109] K. Kyritsis, C. L. Tatli, C. Diou, and A. Delopoulos. Automated analysis of in meal eating behavior using a commercial wristband imu sensor. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2843–2846, July 2017.

[110] Konstantinos Kyritsis, Christos Diou, and Anastasios Delopoulos. Modeling wrist micromovements to measure in-meal eating behavior from inertial sensor data. *IEEE Journal of Biomedical and Health Informatics*, pages 1–1, 2019.

[111] Gierad Laput and Chris Harrison. Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 338:1–338:13, New York, NY, USA, 2019. ACM.

[112] Gierad Laput, Robert Xiao, and Chris Harrison. 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, pages 321–333, New York, NY, USA, 2016. ACM.

[113] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho. Human activity recognition from accelerometer data using convolutional neural network. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 131–134, Feb 2017.

[114] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho. Human activity recognition from accelerometer data using convolutional neural network. In *2017 ieee international conference on big data and smart computing (bigcomp)*, pages 131–134. IEEE, 2017.

[115] Fabrice Lefevre. A dbn-based multi-level stochastic spoken language understanding system. In *2006 IEEE Spoken Language Technology Workshop*, pages 78–81. IEEE, 2006.

[116] Ce Li, Chunyu Xie, Baochang Zhang, Chen Chen, and Jungong Han. Deep fisher discriminant learning for mobile hand gesture recognition. *Pattern Recognition*, 77:276–288, 2018.

[117] Frédéric Li, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors*, 18(2):679, 2018.

[118] Shuai Li, Chuankun Li, Wanqing Li, Yonghong Hou, and Chris Cook. 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 '18, pages 1541–1547, New York, NY, USA, 2018. ACM.

[119] Xi'ang Li, Jinqi Luo, and Rabih Younes. Activitygan: generative adversarial networks for data augmentation in sensor-based human activity recognition. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pages 249–254, 2020.

[120] Yongmou Li, Dianxi Shi, Bo Ding, and Dongbo Liu. Unsupervised feature learning for human activity recognition using smartphone sensors. In Rajendra Prasath, Philip O'Reilly, and T. Kathirvalavakumar, editors, *Mining Intelligence and Knowledge Exploration*, pages 99–107, Cham, 2014. Springer International Publishing.

[121] Yongmou Li, Dianxi Shi, Bo Ding, and Dongbo Liu. Unsupervised feature learning for human activity recognition using smartphone sensors. In *Mining intelligence and knowledge exploration*, pages 99–107. Springer, 2014.

[122] Cihang Liu, Lan Zhang, Zongqian Liu, Kebin Liu, Xiangyang Li, and Yunhao Liu. 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, pages 334–347, New York, NY, USA, 2016. ACM.

[123] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. In *2009 IEEE International Conference on Pervasive Computing and Communications*, pages 1–9, March 2009.

[124] Yun Long, Eui Min Jung, Jaeha Kung, and Saibal Mukhopadhyay. Reram crossbar based recurrent neural network for human activity detection. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 939–946. IEEE, 2016.

[125] Yonghong Luo, Xiangrui Cai, Ying ZHANG, Jun Xu, and Yuan xiaojie. Multivariate time series imputation with generative adversarial networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 1596–1607. Curran Associates, Inc., 2018.

[126] Mingqi Lv, Wei Xu, and Tieming Chen. A hybrid deep convolutional and recurrent neural network for complex activity recognition using multimodal sensors. *Neurocomputing*, 362:33–40, 2019.

[127] Ali Madani, Mehdi Moradi, Alexandros Karargyris, and Tanveer Syeda-Mahmood. Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. In *2018 IEEE 15th International symposium on biomedical imaging (ISBI 2018)*, pages 1038–1042. IEEE, 2018.

[128] Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. Protecting sensory data against sensitive inferences. In *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*, W-P2DS'18, pages 2:1–2:6, New York, NY, USA, 2018. ACM.

[129] Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. Mobile sensor data anonymization. In *Proceedings of the International Conference on Internet of Things Design and Implementation*, IoTDI '19, pages 49–58, New York, NY, USA, 2019. ACM.

[130] Mohammad Malekzadeh, Richard G Clegg, and Hamed Haddadi. Replacement autoencoder: A privacy-preserving algorithm for sensory data analysis. In *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 165–176. IEEE, 2018.

[131] Shinya Matsui, Nakamasa Inoue, Yuko Akagi, Goshu Nagino, and Koichi Shinoda. User adaptation of convolutional neural network for human activity recognition. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 753–757. IEEE, 2017.

[132] Claire Meagher, Enrico Franco, Ruth Turk, Samuel Wilson, Nathan Steadman, Lauren McNicholas, Ravi Vaidyanathan, Jane Burridge, and Maria Stokes. New advances in mechanomyography sensor technology and signal processing: Validity and intrarater reliability of recordings from muscle. *Journal of rehabilitation and assistive technologies engineering*, 7:2055668320916116, 2020.

[133] R. Meattini, S. Benatti, U. Scarcia, D. De Gregorio, L. Benini, and C. Melchiorri. An semg-based human–robot interface for robotic hands using machine learning and synergies. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 8(7):1149–1158, July 2018.

[134] Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10), 2017.

[135] Nastaran Mohammadian Rad, Twan Van Laarhoven, Cesare Furlanello, and Elena Marchiori. Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in parkinson's disease and autism spectrum disorders. *Sensors*, 18(10), 2018.

[136] S. Mohammed and I. Tashev. Unsupervised deep representation learning to remove motion artifacts in free-mode body sensor networks. In *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 183–188, May 2017.

[137] Jafet Morales and David Akopian. Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomedical Engineering*, 37(3):388 – 400, 2017.

[138] Fernando Moya Rueda, René Grzeszick, Gernot A Fink, Sascha Feldhorst, and Michael Ten Hompel. Convolutional neural networks for human activity recognition using body-worn sensors. In *Informatics*, volume 5, page 26. Multidisciplinary Digital Publishing Institute, 2018.

[139] Mario Munoz-Organero and Ramona Ruiz-Blazquez. Time-elastic generative model for acceleration time series in human activity recognition. *Sensors*, 17(2):319, 2017.

[140] Abdulmajid Murad and Jae-Young Pyun. Deep recurrent neural networks for human activity recognition. *Sensors*, 17(11), 2017.

[141] Kouichi Murakami and Hitomi Taguchi. Gesture recognition using recurrent neural networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, page 237–242, New York, NY, USA, 1991. Association for Computing Machinery.

[142] Ronald Mutegeki and Dong Seog Han. A cnn-lstm approach to human activity recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 362–366. IEEE, 2020.

[143] Augustus Odena. Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583*, 2016.

[144] Jeremiah Okai, Stylianos Paraschiakos, Marian Beekman, Arno Knobbe, and Cláudio Rebelo de Sá. Building robust models for human activity recognition from raw accelerometers data using gated recurrent units and long short term memory neural networks. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2486–2491. IEEE, 2019.

[145] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[146] C. Pacchierotti, G. Salvietti, I. Hussain, L. Meli, and D. Prattichizzo. The hring: A wearable haptic device to avoid occlusions in hand tracking. In *2016 IEEE Haptics Symposium (HAPTICS)*, pages 134–139, April 2016.

[147] Nawadita Parajuli, Neethu Sreenivasan, Paolo Bifulco, Mario Cesarelli, Sergio Savino, Vincenzo Niola, Daniele Esposito, Tara J Hamilton, Ganesh R Naik, Upul Gunawardana, et al. Real-time emg based pattern recognition control for hand prostheses: a review on existing methods, challenges and future implementation. *Sensors*, 19(20):4596, 2019.

[148] Liangying Peng, Ling Chen, Zhenan Ye, and Yi Zhang. Aroma: A deep multi-task learning based simple and complex human activity recognition method using wearable sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):74:1–74:16, July 2018.

[149] Angkoon Phinyomark and Erik Scheme. Emg pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing*, 2(3), 2018.

[150] T. Plötz and Y. Guan. Deep learning for human activity recognition in mobile computing. *Computer*, 51(5):50–59, May 2018.

[151] Yanjun Qin, Haiyong Luo, Fang Zhao, Chenxing Wang, Jiaqi Wang, and Yuexia Zhang. Toward transportation mode recognition using deep convolutional and long short-term memory recurrent neural networks. *IEEE Access*, 7:142353–142367, 2019.

[152] Zhen Qin, Yibo Zhang, Shuyu Meng, Zhiguang Qin, and Kim-Kwang Raymond Choo. Imaging and fusing time series for wearable sensor-based human activity recognition. *Information Fusion*, 53:80–87, 2020.

[153] Nastaran Mohammadian Rad, Twan van Laarhoven, Cesare Furlanello, and Elena Marchiori. Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in parkinson's disease and autism spectrum disorders. In *Sensors*, 2018.

[154] Valentin Radu, Nicholas D. Lane, Sourav Bhattacharya, Cecilia Mascolo, Mahesh K. Marina, and Fahim Kawsar. 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, pages 185–188, New York, NY, USA, 2016. ACM.

[155] Mahbubur Rahman, Nasir Ali, Rummana Bari, Nazir Saleheen, Mustafa al'Absi, Emre Ertin, Ashley Kennedy, Kenzie L. Preston, and Santosh Kumar. *mDebugger: Assessing and Diagnosing the Fidelity and Yield of Mobile Sensor Data*, pages 121–143. Springer International Publishing, Cham, 2017.

[156] Giorgia Ramponi, Pavlos Protopapas, Marco Brambilla, and Ryan Janssen. T-CGAN: conditional generative adversarial network for data augmentation in noisy time series with irregular sampling. *CoRR*, abs/1811.08295, 2018.

[157] Daniele Ravi, Charence Wong, Benny Lo, and Guang-Zhong Yang. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*, pages 71–76. IEEE, 2016.

[158] A. Reiss and D. Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th International Symposium on Wearable Computers*, pages 108–109, June 2012.

[159] Meticulous Research. 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. [Online]. Available: `https://www.meticulousresearch.com/product/wearable-devices-market-5050`,. Accessed: Dec 2019.

[160] Patricio Rivera, Edwin Valarezo, Mun-Taik Choi, and Tae-Seong Kim. Recognition of human hand activities based on a single wrist imu using recurrent neural networks. *International Journal of Pharma Medicine and Biological Sciences*, 6(4):114–118, 2017.

[161] Charissa Ann Ronao and Sung-Bae Cho. Deep convolutional neural networks for human activity recognition with smartphone sensors. In *International Conference on Neural Information Processing*, pages 46–53. Springer, 2015.

[162] Charissa Ann Ronao and Sung-Bae Cho. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.*, 59(C):235–244, Oct 2016.

[163] Adrian Rubio-Solis, George Panoutsos, Carlos Beltran-Perez, and Uriel Martinez-Hernandez. A multilayer interval type-2 fuzzy extreme learning machine for the recognition of walking activities and gait events using wearable sensors. *Neurocomputing*, 389:42–55, 2020.

[164] Michael S. Ryoo and Larry Matthies. First-person activity recognition: What are they doing to me? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[165] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. Synthesizing and reconstructing missing sensory modalities in behavioral context recognition. *Sensors*, 18:2967, 09 2018.

[166] Aaqib Saeed, Tanir Ozcelebi, and Johan J. Lukkien. Synthesizing and reconstructing missing sensory modalities in behavioral context recognition. In *Sensors*, 2018.

[167] Oluwarotimi Williams Samuel, Mojisola Grace Asogbon, Yanjuan Geng, Ali H Al-Timemy, Sandeep Pirbhulal, Ning Ji, Shixiong Chen, Peng Fang, and Guanglin Li. Intelligent emg pattern recognition control method for upper-limb multifunctional prostheses: advances, current challenges, and future prospects. *Ieee Access*, 7:10150–10165, 2019.

[168] Sadiq Sani, Stewart Massie, Nirmalie Wiratunga, and Kay Cooper. Learning deep and shallow features for human activity recognition. In *International conference on knowledge science, engineering and management*, pages 469–482. Springer, 2017.

[169] Anton Maximilian Schäfer and Hans Georg Zimmermann. Recurrent neural networks are universal approximators. In Stefanos D. Kollias, Andreas Stafylopatis, Włodzisław Duch, and Erkki Oja, editors, *Artificial Neural Networks – ICANN 2006*, pages 632–640, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[170] Volkan Y Senyurek, Masudul H Imtiaz, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov. A cnn-lstm neural network for recognition of puffing in smoking episodes using wearable sensors. *Biomedical Engineering Letters*, 10(2):195–203, 2020.

[171] W Seok, Youngjoo Kim, and Cheolsoo Park. Pattern recognition of human arm movement using deep reinforcement learning. In *2018 International Conference on Information Networking (ICOIN)*, pages 917–919. IEEE, 2018.

[172] Xiaoran Shi, Yaxin Li, Feng Zhou, and Lei Liu. Human activity recognition based on deep learning method. In *2018 International Conference on Radar (RADAR)*, pages 1–5. IEEE, 2018.

[173] Yunzhou Shi, Megan Manco, Dominique Moyal, Gil Huppert, Hitoshi Araki, Anthony Banks, Hemant Joshi, Richard McKenzie, Alex Seewald, Guy Griffin, Ellora Sen-Gupta, Donald Wright, Philippe Bastien, Florent Valceschini, Sophie Seité, John A. Wright, Roozbeh Ghaffari, John Rogers, Guive Balooch, and Rafal M. Pielak. Soft, stretchable, epidermal sensor with integrated electronics and photochemistry for measuring personal uv exposures. *PLOS ONE*, 13(1):1–15, 01 2018.

[174] S. Shin and W. Sung. Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2274–2277, May 2016.

[175] M. Shoaib, H. Scholten, and P. J. M. Havinga. Towards physical activity recognition using smartphone sensors. In *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, pages 80–87, Dec 2013.

[176] Muhammad Shoaib, Stephan Bosch, Ozlem Incel, Hans Scholten, and Paul Havinga. Fusion of smartphone motion sensors for physical activity recognition. *Sensors*, 14(6):10146–10176, 2014.

[177] Nabeel Siddiqui and Rosa H. M. Chan. A wearable hand gesture recognition device based on acoustic measurements at wrist. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4443–4446, 2017.

[178] Luis Sigcha, Nélson Costa, Ignacio Pavón, Susana Costa, Pedro Arezes, Juan Manuel López, and Guillermo De Arcas. Deep learning approaches for detecting freezing of gait in parkinson's disease patients through on-body acceleration sensors. *Sensors*, 20(7):1895, 2020.

[179] Pekka Siirtola and Juha Röning. Incremental learning to personalize human activity recognition models: The importance of human ai collaboration. *Sensors*, 19(23), 2019.

[180] Elnaz Soleimani and Ehsan Nazerfard. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *CoRR*, abs/1903.12489, 2019.

[181] Jost Tobias Springenberg. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv preprint arXiv:1511.06390*, 2015.

[182] Odongo Steven Eyobu and Dong Seog Han. Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network. *Sensors*, 18(9), 2018.

[183] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, SenSys '15, page 127–140, New York, NY, USA, 2015. Association for Computing Machinery.

[184] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 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, pages 127–140, New York, NY, USA, 2015. ACM.

[185] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 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, page 127–140, New York, NY, USA, 2015. Association for Computing Machinery.

[186] Jian Sun, Yongling Fu, Shengguang Li, Jie He, Cheng Xu, and Lin Tan. Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. *Journal of Sensors*, 2018, 2018.

[187] Subramanian Sundaram, Petr Kellnhofer, Yunzhu Li, Jun-Yan Zhu, Antonio Torralba, and Wojciech Matusik. Learning the signatures of the human grasp using a scalable tactile glove. *Nature*, 569(7758), 2019.

[188] Richard S. Sutton and Andrew G. Barto. Reinforcement learning: An introduction. 2018.

[189] Akira Tamamori, Tomoki Hayashi, Tomoki Toda, and Kazuya Takeda. An investigation of recurrent neural network for daily activity recognition using multi-modal signals. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1334–1340. IEEE, 2017.

[190] Akira Tamamori, Tomoki Hayashi, Tomoki Toda, and Kazuya Takeda. Daily activity recognition based on recurrent neural network using multi-modal signals. *APSIPA Transactions on Signal and Information Processing*, 7, 2018.

[191] Yin Tang, Qi Teng, Lei Zhang, Fuhong Min, and Jun He. Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. *IEEE Sensors Journal*, 21(1):581–592, 2021.

[192] Dapeng Tao, Yonggang Wen, and Richang Hong. Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition. *IEEE Internet of Things Journal*, 3(6):1124–1134, 2016.

[193] Triwiyanto Triwiyanto, I Putu Alit Pawana, and Mauridhi Hery Purnomo. An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(7):1678–1688, 2020.

[194] Md Zia Uddin, Mohammed Mehedi Hassan, Ahmed Alsanad, and Claudio Savaglio. A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. *Information Fusion*, 55:105–115, 2020.

[195] Terry T. Um, Franz M. J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 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 '17, pages 216–220, New York, NY, USA, 2017. ACM.

[196] Muhammad Zia ur Rehman, Asim Waris, Syed Omer Gilani, Mads Jochumsen, Imran Khan Niazi, Mohsin Jamil, Dario Farina, and Ernest Nlandu Kamavuako. Multiday emg-based classification of hand motions with deep learning techniques. *Sensors*, 18(8):2497, 2018.

[197] Y. Vaizman, K. Ellis, and G. Lanckriet. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Computing*, 16(4):62–74, October 2017.

[198] Añazco Edwin Valarezo, Lopez Patricio Rivera, Hyemin Park, Nahyeon Park, and Tae-Seong Kim. Human activities recognition with a single writs imu via a variational autoencoder and android deep recurrent neural nets. *Computer Science and Information Systems*, (00):5–5, 2020.

[199] P. Vamplew and A. Adams. Recognition and anticipation of hand motions using a recurrent neural network. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 6, pages 2904–, 1995.

[200] Tim van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 1–9, New York, NY, USA, 2008. ACM.

[201] Alireza Abedin Varamin, Ehsan Abbasnejad, Qinfeng Shi, Damith C Ranasinghe, and Hamid Rezatofighi. 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*, pages 246–253, 2018.

[202] George Vavoulas, Charikleia Chatzaki, Thodoris Malliotakis, Matthew Pediaditis, and Manolis Tsiknakis. The mobiact dataset: Recognition of activities of daily living using smartphones. In *ICT4AgeingWell*, pages 143–151, 2016.

[203] Emily A. Vogels. About one-in-five Americans use a smart watch or fitness tracker. [Online]. Available: `https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/`,. Accessed: Jan 2020.

[204] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018, 2018.

[205] Chi Cuong Vu and Jooyong Kim. Human motion recognition by textile sensors based on machine learning algorithms. *Sensors*, 18(9):3109, 2018.

[206] Huihui Wang, Lianfu Li, Hao Chen, Yi Li, Sen Qiu, and Raffaele Gravina. Motion recognition for smart sports based on wearable inertial sensors. In *EAI International Conference on Body Area Networks*, pages 114–124. Springer, 2019.

[207] Jiwei Wang, Yiqiang Chen, Yang Gu, Yunlong Xiao, and Haonan Pan. Sensorygans: An effective generative adversarial framework for sensor-based human activity recognition. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[208] Kun Wang, Jun He, and Lei Zhang. Attention-based convolutional neural network for weakly labeled human activities recognition with wearable sensors. *CoRR*, abs/1903.10909, 2019.

[209] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Sami Mekki, Stefan Valentin, and Daniel Roggen. 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 '18, pages 1626–1635, New York, NY, USA, 2018. ACM.

[210] Lukun Wang. Recognition of human activities using continuous autoencoders with wearable sensors. *Sensors*, 16(2):189, 2016.

[211] Xufei Wang, Weixian Liao, Yifan Guo, Lixing Yu, Qianlong Wang, Miao Pan, and Pan Li. Perrnn: Personalized recurrent neural networks for acceleration-based human activity recognition. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2019.

[212] Yucheng Wang, Chunhui Wang, Zhonghui Wang, Xiaojie Wang, and You Li. Hand gesture recognition using sparse autoencoder-based deep neural network based on electromyography measurements. In *Nano-, Bio-, Info-Tech Sensors, and 3D Systems II*, volume 10597, page 105971D. International Society for Optics and Photonics, 2018.

[213] G. M. Weiss, J. W. Lockhart, T. T. Pulickal, P. T. McHugh, I. H. Ronan, and J. L. Timko. Actitracker: A smartphone-based activity recognition system for improving health and well-being. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 682–688, Oct 2016.

[214] Haifeng Wu, Qing Huang, Daqing Wang, and Lifu Gao. A cnn-svm combined model for pattern recognition of knee motion using mechanomyography signals. *Journal of Electromyography and Kinesiology*, 42:136–142, 2018.

[215] Yuheng Wu, Bin Zheng, and Yongting Zhao. Dynamic gesture recognition based on lstm-cnn. In *2018 Chinese Automation Congress (CAC)*, pages 2446–2450. IEEE, 2018.

[216] Kun Xia, Jianguang Huang, and Hanyu Wang. Lstm-cnn architecture for human activity recognition. *IEEE Access*, 8:56855–56866, 2020.

[217] Cheng Xu, Duo Chai, Jie He, Xiaotong Zhang, and Shihong Duan. Innohar: A deep neural network for complex human activity recognition. *IEEE Access*, 7:9893–9902, 2019.

[218] Shuyuan Xu, Jun Wang, Wenchi Shou, Tuan Ngo, Abdul-Manan Sadick, and Xiangyu Wang. Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering*, pages 1–15, 2020.

[219] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[220] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Tröster. 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, pages 17–33, Berlin, Heidelberg, 2008. Springer-Verlag.

[221] Tahmina Zebin, Matthew Sperrin, Niels Peek, and Alexander J Casson. Human activity recognition from inertial sensor time-series using batch normalized deep lstm recurrent networks. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–4. IEEE, 2018.

[222] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services*, pages 197–205, Nov 2014.

[223] Ming Zeng, Haoxiang Gao, Tong Yu, Ole J Mengshoel, Helge Langseth, Ian Lane, and Xiaobing Liu. Understanding and improving recurrent networks for human activity recognition by continuous attention. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pages 56–63, 2018.

[224] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *6th international conference on mobile computing, applications and services*, pages 197–205. IEEE, 2014.

[225] Ming Zeng, Tong Yu, Xiao Wang, Le T. Nguyen, Ole J. Mengshoel, and Ian Lane. Semi-supervised convolutional neural networks for human activity recognition. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 522–529, 2017.

[226] A. Zhang, A. Cebulla, S. Panev, J. Hodgins, and F. De la Torre. Weakly-supervised learning for parkinson's disease tremor detection. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 143–147, July 2017.

[227] Chaolong Zhang, Yigang He, Lifeng Yuan, and Sheng Xiang. Analog circuit incipient fault diagnosis method using dbn based features extraction. *Ieee Access*, 6:23053–23064, 2018.

[228] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, Omer Inan, and Gregory D. Abowd. 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, pages 437:1–437:10, New York, NY, USA, 2018. ACM.

[229] L. Zhang, X. Wu, and D. Luo. Real-time activity recognition on smartphones using deep neural networks. In *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)*, pages 1236–1242, Aug 2015.

[230] L. Zhang, X. Wu, and D. Luo. Recognizing human activities from raw accelerometer data using deep neural networks. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 865–870, Dec 2015.

[231] Mi Zhang and Alexander A. Sawchuk. 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, pages 1036–1043, New York, NY, USA, 2012. ACM.

[232] Xiang Zhang, Lina Yao, Chaoran Huang, Sen Wang, Mingkui Tan, Guodong Long, and Can Wang. Multi-modality sensor data classification with selective attention. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 3111–3117. International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[233] Yun C. Zhang, Shibo Zhang, Miao Liu, Elyse Daly, Samuel Battalio, Santosh Kumar, Bonnie Spring, James M. Rehg, and Nabil Alshurafa. Syncwise: Window induced shift estimation for synchronization of video and accelerometry from wearable sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(3), September 2020.

[234] Z. Zhang, Z. Pi, and B. Liu. Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Biomedical Engineering*, 62(2):522–531, Feb 2015.

[235] Hongyang Zhao, Yongsen Ma, Shuangquan Wang, Amanda Watson, and Gang Zhou. Mobigesture: Mobility-aware hand gesture recognition for healthcare. *Smart Health*, 9-10:129 – 143, 2018. CHASE 2018 Special Issue.

[236] Jianian Zheng, Huiyi Cao, Diliang Chen, Rahila Ansari, Kuo-Chung Chu, and Ming-Chun Huang. Designing deep reinforcement learning systems for musculoskeletal modeling and locomotion analysis using wearable sensor feedback. *IEEE Sensors Journal*, 20(16):9274–9282, 2020.

[237] Longfei Zheng, Shuai Li, Ce Zhu, and Yanbo Gao. Application of indrnn for human activity recognition: The sussex-huawei locomotion-transportation challenge. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 869–872, 2019.

[238] Ding-Xuan Zhou. Universality of deep convolutional neural networks. *Applied and computational harmonic analysis*, 48(2):787–794, 2020.

[239] J. Zhu, A. Pande, P. Mohapatra, and J. J. Han. Using deep learning for energy expenditure estimation with wearable sensors. In *2015 17th International Conference on E-health Networking, Application Services (HealthCom)*, pages 501–506, Oct 2015.

[240] Yida Zhu, Haiyong Luo, Runze Chen, Fang Zhao, and Lin Su. Densenetx and gru for the sussex-huawei locomotion-transportation recognition challenge. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pages 373–377, 2020.